SAPIENZA
UNIVERSITÀ DI ROMA

Use of Next Generation Sequencing for isolated and syndromic Anophthalmia, Microphthalmia and Coloboma (MAC): a new approach to molecular genetic diagnosis

Department of Cellular Biotechnologies and Hematology
PhD in Human Biology and Medical Genetics
XXX° cycle 2014-2017

Candidate
Dr. Iapichino Giuseppe
1235269

Supervisor
Prof. Pizzuti Antonio

Correlator
Prof. Novelli Antonio

A/A 2016/2017

1. INTRODUCTION

1.1 Eye embryogenesis

1.2 MAC: Mixed group of diseases

1.2.1 Microphthalmia, Anophthalmia and Coloboma

1.2.2 Syndromic diseases:

- CHARGE syndrome
- Cornelia de Lange syndrome
- Axenfeld-Rieger syndrome
- Lenz's Microphtalmia
- Papillorenal syndrome

1.3 Molecular Diagnosis: Next Generation Sequencing

2. AIM OF THE STUDY

3. MATERIALS AND METHODS

3.1 Patients Recruitment

3.2 Molecular Analysis

3.2.1 Blood collection, DNA extraction and quantification

3.2.2 NGS panel construction: *Design Studio*

3.2.3 NextSeq-500: features and workflow

- Libraries preparation: Nextera Rapid Capture Custom Enrichment
- Cluster generation
- Sequencing

3.2.4 Sanger Sequencing

3.2.5 Data Analysis

4. RESULTS

5. DISCUSSION

6. CONCLUSION

7. BIBLIOGRAPHY

1. INTRODUCTION

1.1 The eye: embryogenesis

Eye development begins around the 4th week of gestation and can be summarized in four main stages [Zagozewski J.L. et al.; 2014]:

- Optic vesicle formation
- Induction of crystalline lens
- Retinal tissue formation
- Optic fissure closure

Each of these events is highly regulated and coordinated by various transcription factors and circulating molecules [Adler R. et al.; 2007]. The first morphological signs of eye development are two protrusions that form optic vesicle in the lateral walls of the prosencephalon. In this process, at a molecular level, an important role is played by the gene RAX, encoding for a protein that helps ectoderm cells migration and movement, inhibiting the expression of a cell-adhesion factor, Nicam [Bailey T.J. et al.; 2004]. The optic vesicles develop and infold to form the optic cups, which connect to the diencephalus via the optic peduncles.

Meanwhile, the ectoderm, covering each optic cup, is organized to form the lens placode, which thicken giving rise to the crystalline lens. At this stage, the gene PAX6 is very important, because PAX6 protein, together with other transcription factors, such as SOX2, allows the synthesis of the proteins that will form the lens [Shaham O. et al.; 2012]. In this process, optic cup and optic peduncle will form, respectively, retina and optic nerve and several genes are involved in this step, such as SHH and PAX2. SHH promotes the development of optical cup cells to retina. PAX2 is instead involved in the separation between the optic cup cells, which will be part of retinal epithelium, and the pedicle cells, which will form the axon outgoing from the ocular bulb [Chow R.L. et al.; 2011]. At this stage, it is important that the distal portions of the optical cup cells (future retinal cells) come closer, thus allowing the closure of the optic fissure [Adler R. et al.; 2007]. Morphogenesis of the eye is completed when the optic fissure is closed and

retinal neurogenesis is ongoing. Once the formation of crystalline lens and retina is completed, the surrounding mesoderma completes the development, forming the choroid and the sclera [Barishak Y.R.; 1992].

1.2 MAC: Mixed group of diseases

1.2.1 Microphthalmia, Anophthalmia and Coloboma

Anophthalmia, microphthalmia (A/M) and coloboma are eye defects caused by disfunctions during the eye development [Ragge N.K. et al.; 2007]. The "anophthalmia" term is referred to a complete absence of the ocular globe even if the ocular annexes, such as conjunctiva and lacrimal apparatus, are present. The microphthalmia is a condition where the ocular globe, present in the orbit, is smaller than normal [Verma A.S. et al.; 2007 ; Plaisancie J. et al.; 2016]. The parameters used to define the size of eyeball are the corneal diameter and the total axial length (TAL),that represents the distance, in millimeters, from the cornea apex to the posterior part of the ocular bulb. Microphthalmia is classified in severe, simple and nano microphthalmia, based on the reduction of the total axial length. The most severe forms show a corneal diameter <4 mm and a TAL <10 mm at birth, or <12 mm after the first year of age [Bardakjian T. et al.; 2004]. The term "Coloboma" is referred to ocular malformations that often match with microphthalmia and result from failure of closure of the optic fissure during the embryogenesis [Warburg M.; 1991]. It may involve several eye structures, as the iris, choroid, ciliary body, retina and optic nerve. The affected patients present consequences from mild to severe. If the embryonal defect involves only iris, the patient does not report any damage to the vision, but when the failure of closure of the optic fissure involves the optic nerve a more severe coloboma occurs [Cunliffe H.E. et al.; 1998]. These pathologies, enclosed under the acronym of MAC, can be found in isolated forms, with defects interesting exclusively the visual structure, or in syndromic forms, with more complex pathological features. Some of the syndromic forms can show an anophthalmia associated to an esophageal and/or genital anomaly, anophthalmia combined to pituitary anomalies, microphthalmia associated to linear skin defects or

anophthalmia associated to pituitary anomalies in the presence of polysindactily [Pasutto F. et al.; 2007 ; Williamson K.A. et al; 2014].

The combined presence of these conditions is 30/100.000 births, and more detailed statistical studies show a range from 0.6 to 4.2 on 100,000 births for anophthalmia, from 2 to 17 out of 100,000 births for microphthalmia and from 2 to 14 on 100,000 births for coloboma [Morrison D. et al.; 2002]. In addition, in the 11% of blind children is reported microphthalmia. Anophthalmia and microphthalmia have complex aetiology, with chromosomal, monogenic and enviromental causes identified. Chromosomal duplications, deletions and translocations are implicated [Williamson K.A. et al; 2014] and also environmental factors can play a role in the pathogenesis of the disease. The strongest evidence for the environmental influence appears to be associated to gestational-acquired infections, but may also include maternal vitamin A deficiency, exposure to X-rays, solvent abuse and thalidomide exposure [O'Keefe M. et al.; 1987]. Patients are often treated by multidisciplinary teams, including ophthalmologists, pediatricians and clinical geneticists, especially in syndromic cases. Diagnosis is based on clinical examination together to imaging diagnostic examinations, including A-scan ultrasonography, to measure total axial length, B-scan ultrasonography to measure the internal structures of the globe and computerized tomography scan (CT) or MRI (magnetic resonance imaging) of the brain and orbits to evaluate both the size and internal structures of the globe, as well as the optical nerves and the extraocular muscle structure [Albernaz V.S. et al.; 1997]. At a genetic level it is important the genetic counseling to identify the complete pathological picture and the family history, and, of course, the chromosome analysis and molecular genetic tests. Among the genes associated with MACs there are SHH, PAX6, PAX2, RAX, SOX2, GDF6, VSX2, OTX2, HCCS, BMP4. In anophthalmia or severe microphthalmia case, usually the two most frequently mutated genes are SOX2 and OTX2, respectively in 10% and 3% of cases [Schneider A. et al., 2009]. VSX2, PAX6, RAX, BMP4 and GDF6 mutations are more rare [Azuma N. et al.; 1999 ; Henderson R.A. et al; 2007]. Both clinical and experimental evidences suggest that these pathologies have a heterogeneous genetic basis. The resulting phenotypes share some clinical features, and this aspect reflects the complex network involved in organogenesis of the eye, where signal molecules,

transcription factors, down regulation mechanisms of cell cycle components and the inhibition of cell-adhesion factors play an important role.

1.2.2 Syndromic Diseases

- CHARGE Syndrome

CHARGE syndrome is characterized by a combination of phenotypic aspects, where the main features are enclosed in the acronym of its name: coloboma, heart malformation, atresia choanae (choanal atresia), growth retardation, genital abnormalities and ear abnormalities. Subsequently, other abnormalities have been added, such as dysmorphia, romboencephalus dysfunction, hypoplasia of semicircular channels and holoprosencephaly. CHARGE syndrome is a rare disease transmitted as an autosomal dominant tract and its prevalence is 1/10,000 births; its clinical diagnosis is established according to the diagnostic criteria delineated by Blake [Blake K.D. et al.; 2006]. One of the main criteria in order to diagnose CHARGE syndrome is to propose the co-existence of coloboma and choanal atresia. Mutations in CHD7 gene are found in more than 70% of patients diagnosed with this syndrome [Lalani S.R. et al.; 2006]. CHD7 is located on chromosome 8 at position q12.1, it is a large gene (188 kb) which contains 37 exons. Its protein product comprises 2997 amino acids; it belongs to the family of chromodomain helicase DNA-binding protein, which comprises 9 members (CHD1-CHD9). CHD7 protein can form complexes with different proteins or can establish specific bonds with different regions of the genome, allowing temporal and specific-tissue regulation of gene expression. Indeed, its action implies a modification of the chromatin structure, altering the accessibility of the transcription apparatus to DNA. Its role as a regulatory element explains the several phenotypic effects of its mutations [Pisaneschi E. et al.; 2015 ; Bergman J.E. et al.; 2011].

- Cornelia de Lange Syndrome

Cornelia de Lange Syndrome is a clinically heterogeneous disease with an autosomal dominant mode of inheritance. Classic Cornelia de Lange syndrome (CdLS) is characterized by distinctive facial features, growth retardation (prenatal onset; <5th centile throughout life), hirsutism, and upper limb reduction defects that range from subtle phalangeal abnormalities to oligodactyly (missing digits). Craniofacial features include synophrys, highly arched eyebrows, long eyelashes, short nose with anteverted nares, small widely spaced teeth, and microcephaly. IQ ranges from below 30 to 102 (mean: 53). Many patients demonstrate autistic and self-destructive tendencies. Frequent findings include cardiac septal defects, gastrointestinal dysfunction, hearing loss, myopia and cryptorchidism or hypoplastic genitalia [Deardorff M.A. et al.; 2005]. Individuals with a milder phenotype have less severe growth, cognitive, and limb involvement, but often have facial features similar to CdLS phenotype [Boyle MI et al.; 2015].

The prevalence of CdLS is difficult to estimate as individuals with milder features are likely under-recognized. Published estimates for the prevalence range from 1:100,000 [Pearce P.M. et al.; 1967] to as high as 1:10,000 [Opitz J.M.; 1985]. More recent data from the EUROCAT dataset have estimated the prevalence at 1:50,000 for the classic form of CdLS [Barisic et al.; 2008].

Diagnosis is based on clinical findings and/or the identification of a heterozygous pathogenic variant in *NIPBL, RAD21,* or *SMC3* or a hemizyous pathogenic variant in *HDAC8* or *SMC1A* [Mannini L et al.; 2013]. NIPBL gene encodes a highly conserved protein, also known as delangine, which performs functions such as DNA repair or transcription regulation, and is also involved in cohesin protein regulation. Cohesin is a protein complex, made up of four core subunits: SMC1A, SMC3, RAD21 and STAG. This complex plays a key role in chromosome cohesion and segregation of sister chromatids. Recently, it has been confirmed that mutations in cohesin's regulators and in structural components of the complex are responsible of Cornelia de Lange phenotype [Kaur M. et al; 2016]. Mutations in HDAC8 have also been associated to Cornelia de Lange syndrome [Feng L. et al.; 2014]. HDAC8 is an SMC3 deacetylase;

loss of function mutations found in Cornelia de Lange patients result in an increased activity of acetylated SMC3 and an inadequate dissolution of cohesin during prophase and anaphase. HDAC8 mutations are rarer than those found in NIPBL gene, which involve 60% of individuals with Cornelia de Lange. The cohesin components, SMC1A and SMC3, are respectively mutated in 5% and 1% of patients affected by CDLS [Huisman S. et al.; 2017].

- Axenfeld-Rieger Syndrome

Axenfeld-Rieger syndrome (ARS) is a rare autosomal dominant disorder with ocular anterior segment dysgenesis and systemic anomalies. The patients affected by this pathology can display different phenotypes. Clinical signs can be divided into two categories, ocular and non-ocular. Ocular abnormalities include hypoplasia of the iris, with formation of fissures similar to those found in polycoria and corectopia. The most frequent non-ocular signs are skull-facial dysmorphisms, teeth abnormalities, such as microdontia and hypodontia, redundancy of periombelic skin and jaw hypoplasia. The prevalence of this syndrome is 1/200,000 births. [Song W. et al.; 2017]. Diagnosis is performed through ophthalmological and clinical examinations and is supported by genetic testing. Mutations in PITX2 and FOXC1 genes are often found in individuals with Axenfeld Rieger's syndrome. Both genes encode for transcription factors and approximately 40% of patients have mutations in one of the two genes. Generally PITX2 mutations are associated with ocular anomalies, teeth and umbilical abnormalities, while mutations in FOXC1 gene are reported in patients with isolated eye anomalies or whit eye, hearing and heart anomalies [Kelberman D. et al.; 2011].

- Lenz microphthalmia

Lenz microphthalmia syndrome (LMS) is characterized by unilateral or bilateral microphthalmia and/or clinical anophthalmia with malformations of the ears, teeth, hands, skeleton, and urinary system. Microphthalmia is often accompanied by microcornea and glaucoma. Coloboma is present in approximately 60% of microphthalmic eyes with severity ranging from isolated iris coloboma to coloboma of the ciliary body, choroid and optic disk. Ears may be low set, anteverted, posteriorly rotated, simple, cup shaped, or abnormally modeled [Ng D.; 2002]. Hearing loss has been observed. Dental findings include irregularly shaped, missing, or widely spaced teeth. Duplicated thumbs, syndactyly, clinodactyly, camptodactyly, and microcephaly are common, as are narrow/sloping shoulders, underdeveloped clavicles, kyphoscoliosis, exaggerated lumbar lordosis, long cylindric thorax, and webbed neck. Genitourinary anomalies include hypospadias, cryptorchidism, renal hypoplasia/aplasia, and hydroureter. Approximately 60% of affected males have mild-to-severe intellectual disability or developmental delay. The diagnosis of Lenz microphthalmia syndrome is based on clinical findings. Mild simple microphthalmia can be identified by measuring the axial length of the globe with A-scan ultrasonography. *NAA10* (MCOPS1 locus) and *BCOR* (MCOPS2 locus) are the only two genes known to be associated with Lenz microphthalmia syndrome (LMS) [Hilton et al.; 2009].

- Papillorenal syndrome

Renal coloboma (papillorenal syndrome) is an autosomal dominant condition characterized by renal hypodysplasia and abnormalities of the optic nerve. Abnormal renal structure or function is noted in 92% of affected individuals; ophthalmologic abnormalities are observed in 77% of affected individuals who have pathogenic variants in *PAX2*, that encodes a transcription factor involved in the first stages of eye development. Renal abnormalities can be clinically silent in rare individuals. In most individuals, clinically significant renal insufficiency/renal failure is reported. End-stage renal disease requiring renal transplant is not uncommon. Ophthalmologic abnormalities are typically described as optic nerve coloboma or dysplasia [Bower M.A. et al.; 2007].

Iris colobomas have not been reported in any individual with a *PAX2* pathogenic variant. Ophthalmologic abnormalities may significantly impair vision in some individuals, while others have subtle changes only noted after detailed ophthalmologic examination. Additional clinical findings include hearing loss, which is noted in 7% of individuals with identified pathogenic variants in *PAX2*. The prevalence of renal coloboma syndrome is unknown. More than 180 affected individuals with pathogenic variants have been reported worldwide. The number of individuals without pathogenic variants who have classic findings of renal coloboma syndrome is not known. There is no evidence for a significant founder effect in any population [Bower M. et al.; 2012]. A case of papillorenal syndrome was also described in the literature due to a translocation of the PAX2 gene, t(2; 15)(q31; q26) [Aydin B. et al.; 2013].

1.3 Molecular diagnosis: Next Generation Sequencing

The term “genome” was invented in 1920 by geneticist Hans Winkler; it indicates an hybrid between "gene" and "chromosome" and it refers to the genetic information of an organism contained in its DNA sequence. If a DNA is mutated, genetic information may be altered and proteins may no longer be produced in a functional manner. Some mutations are associated to diseases, while others cause phenotypic variations. These mutations, and the information included in human genome can be studied in several different ways and at different levels [Lewis R., 2010]. In a mendelian or monogenic disease, by performing a genetic analysis of the single causative gene it is possible to obtain a sure diagnosis. When the clinical picture is more complex, it is recommended to investigate a wide range of genes which may be involved in the phenotype. In this diagnostic perspective it is useful to develop a technique able to perform an high-throughput gene sequencing in order to find the causative mutation of the pathologic condition.

DNA sequencing is a process that elaborates the order of the nucleotides of a genomic DNA fragment. In 1977 Fred Sanger and his colleagues developed an approach based on high-fidelity DNA replication which was obtained using a DNA polymerase and dideoxynucleotides [Sanger F. et al., 1977]. This concept has provided the base for the development of automated systems based on Sanger sequencing. Sanger sequencing, also known as first-generation sequencing, has been spread all over the world since the 1980's and has been the method chosen for over 25 years. The big limits of the Sanger method are linked to time, cost and resolution. It is still used to confirm the sequence variants, but Next Generation Sequencing (NGS) is now being widely used for large-scale DNA analysis. NGS, unlike conventional Sanger sequencing, allows the sequencing of many fragments in parallel [Rehm H.L. et al.; 2013]. Several NGS systems have been developed by different companies. All these systems, however, have three main steps: DNA preparation and immobilization (“sequence library preparation”), amplification reaction and sequencing reaction.

The introduction of this efficient technology is a step forward in the field of sequencing with great advantages in the comprehension of human genetics and diseases. The possibility to analyze millions of sequencing reactions in parallel at very reduced costs, with shorter processing times and very little amounts of initial DNA, has provided a

formidable boost for the study of rare diseases. [Glenn T.C.; 2011]. The pathologies that benefited the most from NGS are mainly complex syndromes caused by many mutations in different genes, such as anophthalmia and microphthalmia, which can now be diagnosed in shorter times.

2. AIM OF THE STUDY

In these last few years, both mendelian and complex diseases have reached a higher level of accuracy in the analytical process, with a more efficient diagnostic definition and also a more detailed genotype-phenotype correlation. Complex syndromic pathologies, as microphthalmia and anophthalmia, often accompanied by complex clinical pictures, and characterized by a high level of genetic and phenotypic heterogeneity, can now be handled in a faster and thorough manner.

The aim of this project was to carefully investigate the genes involved in microphthalmia, anophthalmia, coloboma and related syndromic diseases using Next Generation Sequencing. This new molecular approach allows to make a massive sequencing of different genomic regions.

In the present study, a panel of different genes involved in the ocular development has been developed to test and analyze simultaneously and rapidly multiple ocular genes of a patient by using the Illumina-Nextseq 500. The method also allows the simultaneous analysis of multiple patients. Each mutation found was confirmed using the classic Sanger sequencing. This analytical approach supports the clinical evaluation in genetic counseling with molecular results, providing a broader spectrum of the genetic background of a patient that can help in the definition of complex phenotypes.

3. MATERIALS AND METHODS

3.1 Patients recruitment

- Casuistry

The cases included in the present study have been recruited among patients'samples belonging to the Medical Genetics Laboratory Unit of the "Ospedale Pediatrico Bambin Gesù" in Rome, from October 2014 to October 2016. During the third year of this PhD project, the results of the analysis performed on the samples previously recruited were elaborated. The ethnic distribution of selected patients reflects that of the general population, mainly Caucasian in origin from the italian territory. A total of 134 patients were included in the study group, including 41 with microphthalmia/anophthalmia, 34 with CHARGE syndrome, 37 with Cornelia de Lange syndrome and a total of 22 patients with combined diagnosis of Rieger syndrome, Lenz microphthalmia and papillorenal syndrome.

- Ethical evaluation

This study and each activity has been performed in accordance to the protocol of the Helsinki Declaration. The study required the use of human biological material, in particular peripheral blood taken from patients with microphthalmia, anophthalmia and coloboma or rare syndromes in which the ocular defect is associated to other clinical features. Both preventive and operational procedures relating to this project have been addressed to the principles of the Universal Declaration of Bioethics and Human Rights adopted unanimously in 2005 by UNESCO member countries. In addition, a "prior, free, explicit and informed consent of the person involved" was obtained by each subject included in the study or by both parents if the proband was underage at the time of blood sampling, as guaranteed by the Italian ordinance and in accordance to Art. 6 of the Universal Declaration of Bioethics and Human Rights and also in accordance to the "a" and "d" paragraphs of article 6 of the International Declaration on Human Genetic Data. In case of physical or mental impairment of a patient, unable to give his informed consent, the subscription was obtained by the legal tutor. Finally, in accordance to art. 9

"the privacy of the people involved and the confidentiality of their personal data" has been respected. The laboratory has implemented the principles of "Good Laboratory Practices" for the safety of operators and personnel involved, both for the use of chemical reagents and for the handling of biological materials and derived products, with the aim of avoiding the risk of species-specific transmission of viral and not-viral pathogens. We acted according to the principles contained in art. 20 of the Universal Declaration of Bioethics and Human Rights ("Risk Assessment and Management") for an appropriate assessment and proper risk management, minimizing the possibility of error and managing the risk probability with appropriate procedures. The operator behavior was inspired to art.18 of the Universal Declaration of Bioethics and Human Rights. We acted following the criteria of "professionalism, honesty, integrity and transparency". We also declare that there were no conflicts of interest.

3.2 Molecular Analysis

3.2.1 Blood collection, DNA extraction and quantification

- Blood collection

After filling the informed consent the patients underwent blood sampling (5 ml of venous blood, in test tubes containing EDTA). The samples were then recorded, associated to a barcode and carefully stored at -20 ° C.

- DNA Extraction

DNA extraction by peripheral blood lymphocytes was performed through "*QIAamp DNA Blood Mini Kit" (Qiagen)*. This protocol is based on three main steps: a purification step, in which DNA is absorbed for affinity on in silico columns, a series of washing steps with appropriate buffers to eliminate RNA and proteins, and an elution step to collect purified genomic DNA. In detail, 20 µl of 15.1mg/µl proteinase K solution (600AU/ml) and 200 µl of lysis buffer (AL) are added to 200 µl of blood. An incubation phase at 56°C for 10' follows, to allow the lysis of white blood cells. The

cellular lysis causes the leakage of nuclear DNA by dissociation of membrane phospholipids and hydrolysis of cytoplasmic proteins. After the incubation phase at 56°C, 200 µl of 100% ethanol are added. The resulting solution is transferred in a small column containing a silica membrane with high affinity for DNA. The column is then centrifuged at 8000 rpm for 1'. After centrifugation, the columns are placed on new tubes and the filtrate is discarded. After the addition of 500 µl of buffer AW1 (wash buffer 1), used to wash the column filters, a centrifugation at 8000 rpm for 1' is performed. The flowthrough is then removed and the samples are washed with 500 µl of AW2 buffer (wash buffer 2) and centrifuged at 14,000 rpm for 3 min. After placing the columns on new 1.5 ml eppendorf collection tubes, the genomic DNA ligated to the filter is eluted with DNAse free water. This procedure was performed both manually and robotically using Qiagen automatic extractors (QiaCube and QiaSymphony).

- DNA quantification

DNA was quantified both by NanoDrop 1000 spectrophotometer (*ThermoScientific*) and by PicoDrop fluorometer (*PicoDrop Technology*). In spectrophotometric measurements, in addition to the concentration value, it is possible to obtain the ratio between the two optical densities (O.D.260nm / O.D.280nm), which represent a measure of the purity of the extracted sample. Values ranging from 1.8 to 2.0 are representative of high quality samples. An additional analysis to verify the concentration of DNA samples was made using FLx800 Fluorescence Microplate Reader with the graphic support of BioTek's Gen5^{TM} 2.0 Data Analysis software. The fluorescence signal was obtained using the Quant-iT^{TM} PicoGreen kit, a highly sensitive fluorescent dye, useful for double-helix DNA quantification (dsDNA). While methods based on absorbance measurement detect, in addition to dsDNA, also contaminants such as nucleotides, single-stranded nucleic acids, RNAs and proteins, highlighting interferences in data measurement, dyes such as PicoGreen Hoechst (Bisbenzimide) bypass many of these problems. The *PicoGreen*® reagent comes in a concentrated anhydrous dimethylsulfoxide (DMSO) solution. A 1:200 dilution of PicoGreen® was prepared using a 2X solution with 1X TE (10 mM Tris-HCl, 1 mM EDTA, pH 7.5). After creating a calibration curve with standard DNA, the extracted DNA is diluted 1:100 with 1X TE, in a final volume of 100 µl, to which an equal amount of diluted

solution of 2X PicoGreen® Reagent is added. The solution is incubated for 5 minutes protected from light, as the reagent is photosensitive, and then transferred on the plate for analysis with fluorimeter. DNA thus obtained is stored at -20° C, with an average concentration of 25ng/µl.

3.2.2 NGS panel construction: *DesignStudio*

Next Generation Sequencing offers the possibility to make the sequencing of multiple genes in a single experiment. Once the genes to include have been chosen, the appropriate primers or specific probes complementary to the target genomic regions are designed. *DesignStudio* is an *on line software* provided by Illumina that, through different bioinformatic tools, allows the optimal construction of the amplicons or the oligonucleotide probes complementary to the genomic region of interest (*target*). The algorithms in the software work to define a series of factors useful to optimize the panel construction. For example, information on the presence of GC-rich regions, on the specificity of the oligonucleotide probes or on the coverage are obtained. In order to maximize the amplification reaction, optimal size ranges of the amplicons are chosen depending on the starting material, e.g. tissue included in paraffin or blood. Illumina also offers a variety of analytical approach methods, which may involve the use of amplicons or probes, providing different types of chemistry of genomic sequencing, such as Truseq Custom Amplicon (TSCA) kits or Nextera kits Rapid Capture Custom Enrichment. However, the specific technique is chosen according to the laboratory needs. In our laboratory, Nextera kits have been chosen based on probes targeting our genomic regions of interest. After oligonucleotide probes design, panel realization is confirmed and in a short time, about 40 days, the reagent kit is obtained. During a NextSeq run, fragments of the target regions are sequenced. Two different panels were developed over time. The first panel designed included all A/M genes, including also those involved in syndromic conditions (Fig. 1). Subsequently, in order to enhance the experiments performance we decided to design two different panels, one comprising only the A/M genes, and the other with genes involved in syndromic pathologies (Fig.2).

Fig.1 First panel designed comprising MAC and syndromic diseases

GENES	Chromosome Position	Exons	Mut. Freq.	A/D	Phenotypic Characteristic /Syndromes
PAX6	Chr11:31,806,290 – 31,839,559	16		D	aniridia, anophthalmia
SOX2	Chr3:181,429,672 – 181,432,273	1	(4) 10-20%	D	A/M (Microphthalmia, syndromic3)
OTX2	Chr14:57,267,375 – 57,277,234	7	3,3-8%	D	A/M, pituitary abnormalities (microphthalmia, syndromic 5)
VSX2	Chr14:74,706,125 – 74,729,491	5	2%	R	Microphthalmia, cataract, coloboma (microphthalmia, isolated 2)(microphthalmia with coloboma 3)
GDF6	Chr8:97,154,508 – 97,173,070	2	8%	D	A/M, coloboma (microphthalmia, isolated 4)
SHH	Chr7:155,595,508 – 155,605,017	3		D	Colobomatous microphthalmia, anophthalmia, (holoprosencephaly 3)(microphthalmia with coloboma 5)
RAX	Chr18:56,934,217 – 56,940,675	3	2%	R	Anophthalmia, microphthalmia (microphthalmia, isolated 3)
PAX2	Chr10:102,505,418 – 102,589,745	13		D	Papillorenal syndrome
BCOR	ChrX:39,910,449 – 40,036,632	17		X-R	Lenz microphthalmia syndrome (oculofaciocardiodental)(microphthalmia, syndromic 2)
CHD7	Chr8:61,591,289 – 61,779,515	38		D	CHARGE syndrome (microphthalmia, coloboma)
NIPBL	Chr5:36,676,811 – 37,065,971	48	50%	D	Cornelia de Lange syndrome
SMC3	Chr10:112,327,399 – 112,364,442	29	2,50%	D	Cornelia de Lange syndrome
SMC1A	ChrX:53,401,020 – 53,449,668	25	2,50%	X-D	Cornelia de Lange syndrome
HDAC8	ChrX:71,549,316 – 71,793,003	14		X-R	Cornelia de Lange syndrome
RAD21	Chr8:117,858,123 – 117,887,155	14		X-R	Cornelia de Lange syndrome
NDP	ChrX:43,807,974 – 43,832,971	3		X-R	Norrie disease
HCCS	ChrX:11,129,356 – 11,141,254	9		X-D	MIDAS: microphthalmia,dermal aplasia and sclerocornea (microphthalmia, syndromic 7)
PITX2	Chr4:111,538,530 – 111,558,558	6		D	Rieger syndrome
FOXC1	Chr6:1,610,631 – 1,614,179	1		D	Rieger syndrome
ERCC6	Chr10:50,664,441 – 50,747,197	21		R	Cerebrooculofacioskeletal syndrome 1
SIX6	Chr14:60,975,888 - 60,978,575	2		D	Microphthalmia, cataract, nystagmus (microphthalmia with cataract 2)
HESX1	Chr3:57,231,894 – 57,234,330	4		D	Septooptic dysplasia
SIX3	Chr2:45,168,987 – 45,173,266	2		D	Microphthalmia, coloboma, PHPV (holoprosencephaly 2)
STRA6	Chr15:74,471,758 – 74,502,096	26		R	Matthew-Wood syndrome/PDAC syndrome (microphthalmia, syndromic 9)(microphthalmia, isolated, with coloboma 8)
MFRP	Chr11:119,209,602 – 119,217,433	15		R	microphthalmia, isolated 5(nanophthalmos)
GDF3	Chr12:7,842,331 – 7,848,410	2		D	Microphthalmia, isolated 7
FOXE3	Chr1:47,881,694 – 47,883,774	1		R/D	Cataracts, glaucoma
CRYBA4	Chr22:27,017,878 – 27,026,656	6		D	Microphthalmia
BMP4	Chr14:54,416,405 – 54,423,604	6		D	A/M; r (microphthalmia, syndromic 6)
RSK3	Chr6:166,822,853 – 167,275,770	21			

Fig.2 Development of two different panels: one for MAC genes (Fig 2a); the other for syndromic pathologies (Fig 2b)

2a

GENES	Chromosome Position	Exons	Mut. Freq.	A/D	Phenotypic Characteristic /Syndromes
PAX6	Chr11:31,806,290 – 31,839,559	16		D	aniridia, anophthalmia
SOX2	Chr3:181,429,672 – 181,432,273	1	(4) 10-20%	D	A/M (Microphthalmia, syndromic3)
OTX2	Chr14:57,267,375 – 57,277,234	7	3,3-8%	D	A/M, pituitary abnormalities (microphthalmia, syndromic 5)
VSX2	Chr14:74,706,125 – 74,729,491	5	2%	R	Microphthalmia, cataract, coloboma (microphthalmia, isolated 2)(microphthalmia with coloboma 3)
GDF6	Chr8:97,154,508 – 97,173,070	2	8%	D	A/M, coloboma (microphthalmia, isolated 4)
SHH	Chr7:155,595,508 – 155,605,017	3		D	Colobomatous microphthalmia, anophthalmia, (holoprosencephaly 3)(microphthalmia with coloboma 5)
RAX	Chr18:56,934,217 – 56,940,675	3	2%	R	Anophthalmia, microphthalmia (microphthalmia, isolated 3)
SIX6	Chr14:60,975,888 - 60,978,575	2		D	Microphthalmia, cataract, nystagmus (microphthalmia with cataract 2)
HESX1	Chr3:57,231,894 – 57,234,330	4		D	Septooptic dysplasia
SIX3	Chr2:45,168,987 – 45,173,266	2		D	Microphthalmia, coloboma, PHPV (holoprosencephaly 2)
STRA6	Chr15:74,471,758 – 74,502,096	26		R	Matthew-Wood syndrome/PDAC syndrome (microphthalmia, syndromic 9)(microphthalmia, isolated, with coloboma 8)
MFRP	Chr11:119,209,602 – 119,217,433	15		R	microphthalmia, isolated 5(nanophthalmos)
GDF3	Chr12:7,842,331 – 7,848,410	2		D	Microphthalmia, isolated 7
FOXE3	Chr1:47,881,694 – 47,883,774	1		R/D	Cataracts, glaucoma
CRYBA4	Chr22:27,017,878 – 27,026,656	6		D	Microphthalmia
BMP4	Chr14:54,416,405 – 54,423,604	6		D	A/M; r (microphthalmia, syndromic 6)
RSK3/RPS6KA2	Chr6:166,822,853 – 167,275,770	21			

2b

GENES	Chromosome Position	Exons	Mut. Freq.	A/D	Phenotypic Characteristic /Syndromes
PAX2	Chr10:102,505,418 – 102,589,745	13		D	Papillorenal syndrome
BCOR	ChrX:39,910,449 – 40,036,632	17		X-R	Lenz microphthalmia syndrome (oculofaciocardiodental)(microphthalmia, syndromic 2)
CHD7	Chr8:61,591,289 – 61,779,515	38		D	CHARGE syndrome (microphthalmia, coloboma)
NIPBL	Chr5:36,676,811 – 37,065,971	48	50%	D	Cornelia de Lange syndrome
SMC3	Chr10:112,327,399 – 112,364,442	29	2,50%	D	Cornelia de Lange syndrome
SMC1A	ChrX:53,401,020 – 53,449,668	25	2,50%	X-D	Cornelia de Lange syndrome

HDAC8	ChrX:71,549,316 – 71,793,003	14	X-R	Cornelia de Lange syndrome
RAD21	Chr8:117,858,123 – 117,887,155	14	X-R	Cornelia de Lange syndrome
NDP	ChrX:43,807,974 – 43,832,971	3	X-R	Norrie disease
HCCS	ChrX:11,129,356 – 11,141,254	9	X-D	MIDAS: microphthalmia,dermal aplasia and sclerocornea (microphthalmia, syndromic 7)
PITX2	Chr4:111,538,530 – 111,558,558	6	D	Rieger syndrome
FOXC1	Chr6:1,610,631 – 1,614,179	1	D	Rieger syndrome
ERCC6	Chr10:50,664,441 – 50,747,197	21	R	Cerebrooculofacioskeletal syndrome 1
OCRL	ChrX:128,674,252 – 128,726,530	24	X-R	Oculocerebrorenal syndrome (Lowe)

3.2.3 NextSeq-500: features and workflow

The Illumina NextSeq 500 system combines the power of high-throughput sequencing with the simplicity of a desktop sequencer. Its fast, integrated, sample-to-result workflow enables rapid sequencing of exomes, whole genomes, targeted panels, transcriptomes in a single run, with the flexibility to switch to lower-throughput as needed. It also supports “Truseq” and “Nextera” library kits. Flow cells are available in configurations for high output and mid output. Each flow cell type is kitted with a compatible prefilled reagent cartridge. Integrated analysis software performs on-instrument data analysis, which includes image analysis and base calling. The NextSeq uses an implementation of Real-Time Analysis (RTA) called RTA v2, which includes important architecture and feature difference. The sequencing workflow is integrated with BaseSpace, the Illumina genomics computing environment for data analysis, storage, and collaboration. For instrument configured for BaseSpace, library information and run parameters are specified on the BaseSpace Prep Tab. Runs that were set up in BaseSpace appear on the instrument interface during run setup. As the run progresses, output files are streamed in real time to BaseSpace or BaseSpace Onsite. The NextSeq Series enables researchers to keep pace with technology, putting them in control of their sequencing projects. The flexible NextSeq Series enables researchers to switch quickly from one application to another and configure output based on sample volume and coverage needs. Now, even the smallest laboratory can perform any combination of sequencing applications to advance their studies, supporting faster publication of landmark research. [https://support.illumina.com/content/dam/illumina-

support/documents/documentation/system_documentation/nextseq/nextseq-500-system-guide-15046563-02.pdf]

- Libraries preparation: Nextera Rapid Capture Custom Enrichment

Nextera Rapid Capture Custom Enrichment is an all-in-one assay for library preparation and custom target enrichment. Nextera tagmentation coupled with optimized target capture ensures the fastest enrichment workflow time for custom content. The flexible, fully customizable design accommodates up to 15 Mb of custom content so you can focus on the regions of the genome that you care about. Nextera-based library preparation generates adapter-tagged libraries from 50 ng input genomic DNA. Nextera tagmentation of DNA simultaneously fragments and tags DNA without the need for mechanical shearing, using the Nextera transposone that bind tightly to DNA ends, fragmenting and tagging DNA with adapter sequences in a single step, adding by PCR step Index 1 (i7) and Index 2 (i5), the common adapters required for cluster generation and sequencing. Integrated sample barcodes allow the pooling of up to 12 of these adapter ligated sample libraries into a single, hybridization-based, pulldown reaction (Fig. 3A). The pooled libraries are then denatured into single-stranded DNA (Fig. 3B) and biotin-labeled probes complementary to the targeted region are used for the Rapid Capture hybridization (Fig. 3C). Streptavidin beads are added, which bind to the biotinylated probes that are hybridized to the targeted regions of interest (Fig. 3D). Magnetic pulldown of the streptavidin beads enriches the targeted regions that are hybridized to biotinylated probes. (Fig. 3E). The enriched DNA fragments are then eluted from the beads and a second round of Rapid Capture is completed to increase enrichment specificity.

The entire process is completed in only 1.5 days, enabling a single researcher to process up to 12 samples efficiently and at one time all without automation [https://www.illumina.com/content/dam/illuminamarketing/documents/products/datasheets/datasheet_nextera_rapid_capture_custom_enrichment.pdf].

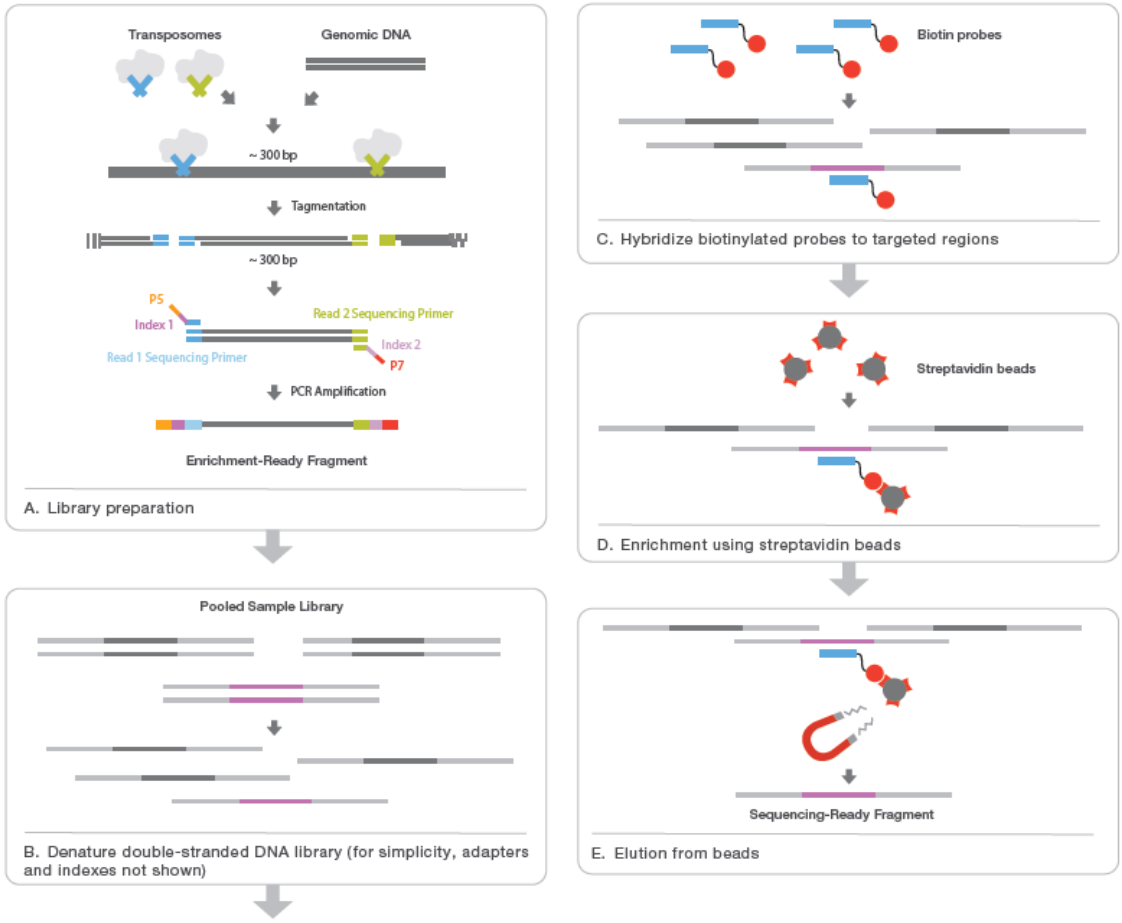

Fig. 3 Nextera Rapid Costum Workflow

- Cluster generation

To perform a sequencing run on the NextSeq 500 we have to insert the reagent cartdrige and a *flow cell* in the instrument. *Flow cell* is a glass-based substrate on which clusters are generated and the sequencing reaction is performed. The flow cell is made of a thin channel divided in *tiles*. Inside the tiles are linked some oligonucleotides, complementary to adapters P5 and P7, which are bound to the end of the sequencing templates. Thanks to the hybridization of the adapters with flow cell oligonucleotides, template stability is guaranteed on the support. Subsequently, libraries are amplified on the surface of this special slide, through a reaction called *bridge amplification* (Fig. 4). This step allows the amplification of up to 1000 identical copies of each molecule enclosed in a cluster, in a space of 1 micron.

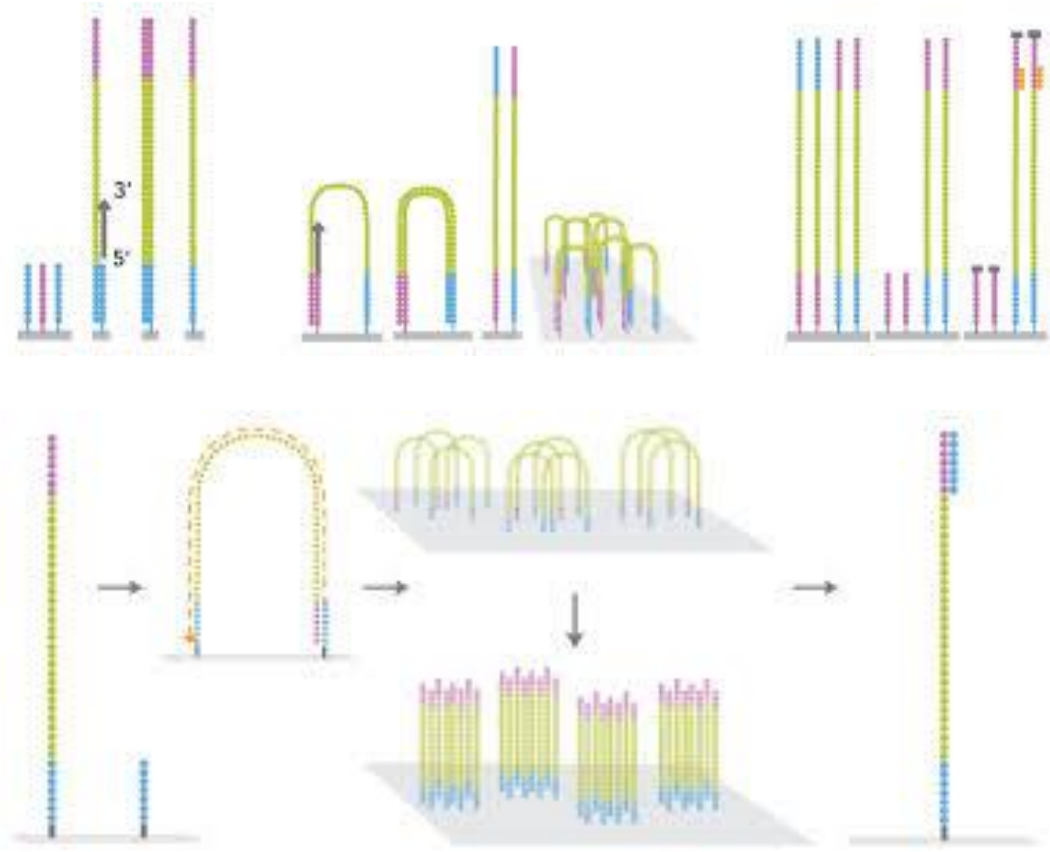

Fig. 4 Cluster generation through Bridge amplification

- Sequencing

Illumina sequencing methodology is defined *Sequencing By Synthesis (SBS)*. Base readings are made in real time. Each sequencing cycle has four steps: incorporation, fluorescence, image and washing (Fig. 5). The four nucleotides are labeled with different fluorocromes, making image capturing possible. Unlike Sanger method, where labeled nucleotides block the binding of the next nucleotide, those used in sequencing by synthesis allow the extension of the new filament. These nucleotides, represented by 3'_O_Ozydomethyl molecules, reproduce reversible terminators. Following the incorporation of a complementary base in the strand of new synthesis, the molecule is excited by the laser and a light signal is registered as image to identify the added base. Subsequently, the 3'OH of the modified nucleotide is restored in order to be able to accommodate a new base. This reaction occurs simultaneously for all filaments

anchored within the flow cell and the images are collected independently for each incorporation. The running performance can be followed in real time using the Sequencing Analysis Viewer (SAV) software. This software provides information on the number and quality of the clusters generated and on the general performance of the experiment.

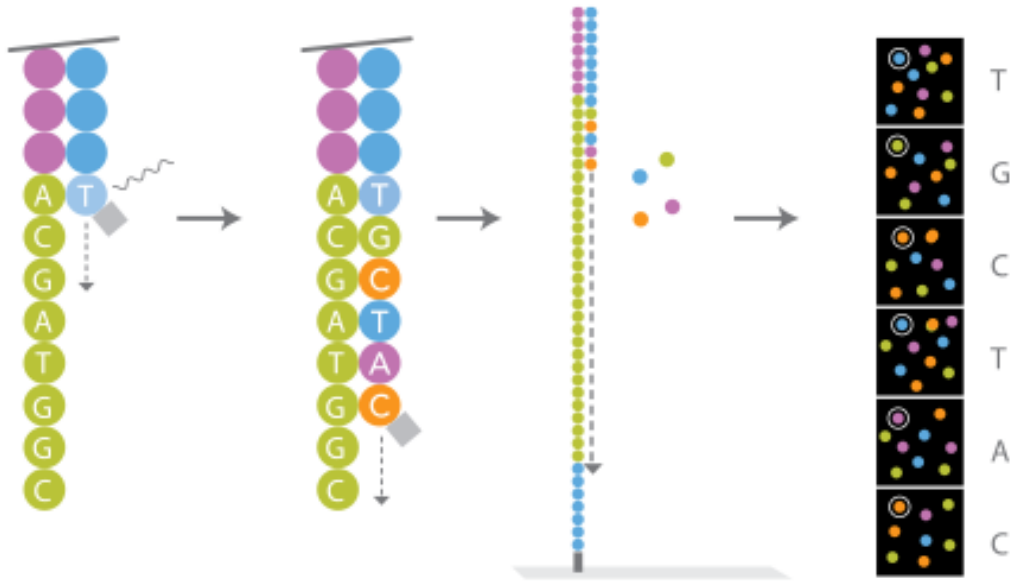

Fig. 5 Sequencing by synthesis

3.2.4 Sanger Sequencing

All mutations found with Next Generation Sequencing have been confirmed by traditional Sanger sequencing, which is also used to analyze genomic regions not covered by NGS sequencing.

- Acquisition of genomic sequences

Coding sequences of genes included in our study have been obtained from online genetic banks, as the Santa Cruz California University Web site (UCSC Genome Bioinformatics- http://genome.ucsc.edu/).

- Primers design

Primers were designed to obtain a maximum amplified length of about 500 bp. For this purpose the Primer3 online software was used, allowing oligonucleotide choice based on the annealing temperature and length. Once a potential pair of primers has been processed by the program, its specificity for the genomic region of interest was checked using Primer BLAST (NCBI) database and SNPcheck online software is interrogated in order to avoid the presence of SNPs in the primer sequences

- Polymerase chain reaction (PCR)

For each PCR reaction about 50ng of genomic DNA, 1pmol/µl of each specific primer for the target sequence, 0,175mM of dNTPs, dimethyl sulfoxide (DMSO) in appropriate quantities, 0,025 unit/µl of DNA Polymerase Kapa Fast (Kapabiosystem, Wilmington, USA) and 1X reaction buffer (Kapa2G Buffer A with Mg) were used. Amplification reactions were performed on Veriti Thermal Cycler (Applied Biosystems). With this type of thermocycler it is possible to amplify samples at different temperatures in the same experiment. This is useful to evaluate the success of the reaction at different annealing temperatures in the same session. PCR reactions start with an enzymatic activation step at 95°C for 1' followed by 30 cycles of: denaturation step at 95°C for 10', annealing step at a specific temperature for each primer pair for 10" and extension step at 72°C for 1". A final extension at 72°C for 30' ends the reaction.

- PCR purification

All components that are not incorporated in the amplification reaction such as salts, primers or primer dimers that may subsequently interfere with the sequencing reaction, are removed by an enzymatic purification with *Illustra ExoproSTAR* method (GE Healthcare Life Sciences). This reagent is composed by two hydrolytic enzymes, an exonuclease I (EXO I), which eliminates primers in excess and single-stranded DNA,

and an alkaline phosphatase (Shrimp Alkaline Phosphatase) that removes dNTPs not incorporated in the PCR reaction hydrolyzing them in nucleosides and inorganic phosphate. 4 µl of PCR product and 1 µl of ExoproSTAR enzyme are added in a PCR tube and inserted in the thermocycler with the following program:

- 37°C for 15' (enzymatic activation)
- 80°C for 15' (enzymatic inactivation)

➢ Sequencing reaction

Purified PCR products can now be sequenced. In Sanger sequencing the sequence is generated through the formation of fragments of different sizes exploiting a controlled interruption of the enzymatic replication. Together with the four deoxyribonucleotides triphosphate (dATP, dTTP, dCTP, dGTP), 2',3'-dideoxynucleotides of each base (ddATP, ddTTP, ddCTP, ddGTP), labeled with different fluorochromes, are added to purified PCR products allowing the extension of the strand based on the amplified template strand. When the dideoxynucleotides are randomly incorporated in the sequence of new synthesis the elongation is interrupted, as the modified nucleotides miss the hydroxyl terminal 3', essential to form the next phosphodiester linkage. Fragments of different lengths with labeled 3' ends are hence casually generated. *BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystem)* is used for sequencing reaction. Two different reactions are prepared for each primer, forward and reverse. Each mix is composed of 1 µl of purified PCR product, 1 µl of primer, 1 µl of BigDye, 1 µl of Buffer BigDye and 6 µl of H_2O to a final volume of 10 µl. The reaction is performed on Veriti Thermal Cycler (Applied Biosystems) and comprises 25 cycles of: denaturation step at 96°C for 10', annealing step at 50°C for 5' and extension step at 60°C for 4'.

- Sequence reaction purification

A new purification step follows the sequencing reaction to eliminate primers and non-incorporated fluorescent nucleotides that would interfere with reading on the automatic capillary sequencer. Different kits are used for purification depending on the amount of samples, single columns of DyeEx 2.0 Spin Kit (QIAGEN), and plates of 96 samples (CENTRISEP - Princeton Separations). Each kit is based on the same principle of chromatography for gel filtration, in which the molecules are separated on the basis of their molecular weight. Purification takes place through a first centrifugation step that removes the liquid in which the resin is stored and making the resin packed into a semi-solid structure. The sequencing reaction is then transferred in the column with the resin and a second centrifugation step is performed. The resin retains the non-incorporated components of the sequencing reaction and the purified PCR product is eluted.

- Sanger Sequencing with automatic capillary electrophoresis

Once the sequencing product has been purified, it is loaded on the 16 capillaries automatic sequencer ABI 3130 or in the 24 capillaries automatic sequencer ABI 3500xL (Applied Biosystem). Within each capillary, each fluorescent nucleotide is electrophoretically separated using POP7 polymer. The polymer's matrix separates fragments based on their size through the potential difference applied by the instrument to both ends of each capillary. As the single fluorescent nucleotide is scanned by the laser, the fluorochrome is excited releasing a fluorescent signal. The intensity and wavelength of the emitted light are captured and measured. The information regarding the color and type of ddNTP present at the end of each filament of different sizes is processed in a software. The information is thus integrated and transformed in color peaks, with areas proportional to the emission intensity. The classic electropherogram is thus formed. Raw sequencing data are analyzed through the Sequencing Analysis program (Applied Biosystem). The program also provides an overview of the quality of the experiment, while Mutation Surveyor software (SoftGenetics) is used for data analysis.

3.2.5 Data Analysis

- *BaseSpace*

Sequencing run data can be processed through a wide range of open-source or commercial pipelines developed for Illumina data, or instantly can be transferred, analyzed, and stored securely in *BaseSpace* (Cloud or Onsite), the Illumina genomics computing environment. BaseSpace downstream data analysis includes alignment and variant detection, annotation, visualization, and interpretation. There is an initial *Demultiplexing phase*, that allows detection of clusters positions on the *flow cell*, needed to discriminate the samples; subsequently the following files are generated:

- *FASTQ_text files*, which contain information on the order of bases in the sequence and reading quality for each nucleotide (*Phred Quality Score_Q).*
- BAM_files are produced by compression of larger SAM files and contain information on alignment of a determined sample compared to a reference sequence and chromosome coordinates.
- VCF_variant call files contain information on variants found in specific positions.

- Variant Studio

All files generated by NextSeq 500 can be loaded on different analysis software to view the variants and the alignment files of the sequence products.
One such software is *Variant Studio*, released by Illumina. [https://www.illumina.com/content/dam/illuminamarketing/documents/products/datasheets/datasheet_illumina_variantstudio_software.pdf]. The software allows to visualize quickly VCF files containing a set of information on the variants found, including gene, variant nomenclature at nucleotide and aminoacid level, the nomenclature based on dbSNP (if described), Sift and PolyPhen data, allelic frequency, variant frequency in the general population and in different populations, the quality of base call, the GQX, the number of reads, the identification of the transcript, exonic or intronic position, classification, coordinates of chromosomic position, if it is a conserved base, the associated pathology if described, links to Google Scholar, PubMed, UCSC Browser, ClinVar RS, ClinVar, OMIM, Orphanet, GeneReviews and many other useful information for the final evaluation of the variants found.

- IGV

Sequence raw data can be viewed through IGV *(Integrative Genomics Viewer)*, a free software downloadable from the Broad Institute site. This interface graphically shows the sequence alignments and the reads (coverage and BAM files) of all sequenced genes, and provides information regarding all variants found (VCF files). The program gives a general idea on the quality of the experiment in toto; it allows to view all the regions not covered by sequencing, regions with little coverage or with low quality of reads. Through IGV we can estimate which regions need to be sequenced through Sanger sequencing to complete the analysis of entire coding regions. [Thorvaldsdóttir H et al.; 2013].

- *Mutation Surveyor*

Exons with probably or definitely pathogenetic variants and those insufficiently covered or completely uncovered are analyzed by Sanger sequencing. Following PCR and classic sequencing reactions of the regions of interest, fragments are analyzed on automatic sequencers ABI3130 XL and ABI3500xL *(Applied Biosystems)*; sequences obtained can then be analyzed with different programs, including Mutation Surveyor (SoftGenetics). Sequences are loaded on the program with the reference sequence (downloadable from GeneBank, in gbk format). Mutation Surveyor identifies the variants by comparing the physical differences between the electrophoretic peaks of the reference and of the test sample sequences [Dong C. et al.; 2013].

- *Alamut*

A tools that helps us to understand the potential pathogenicity of a variant is *Alamut*, an in silico prediction software (Interactive Biosoftware). Alamut is an interactive software that allows to view the gene of interest and to have all the information about it; once the variant has been highlighted all information on the variant are obtained through links to all the online analysis databases. Very important is the evaluation on canonical and non-canonical splicing sites performed by the software. It also provides the level of conservation of the nucleotide and amino acid in various species. It cannot, of course, replace an in vivo study of the variant, but it definitely helps to understand if the variant is to be taken into account [http://www.interactive-biosoftware.com/doc/alamut-batch/1.7.0/EN/Alamut-Batch-1.7.0-User-Manual.pdf].

4. RESULTS

In the present study we analyzed 134 patients, differently distributed among the various pathologies under investigation. In detail, 41 patients with anophthalmia, microphthalmia and Coloboma have been included in this study (Tab. 1). Of these, 12 (29%) displayed variants in the analyzed genes (Tab. 2). Among 34 patients with CHARGE syndrome (Tab. 3), 19 (55%) displayed variants in the CHD7 gene (Tab. 4), while of 37 patients with Cornelia De Lange syndrome (Tab. 5), 7 (19%) showed variants in the associated genes (Tab. 6). Eight patients with Axenfeld-Rieger syndrome have been analyzed (Tab.7). Of these, 2 patients (25%) had mutations in two genes involved in the pathology (Tab. 8). Respectively 8 and 6 patients have syndromes associated to PAX2 and BCOR genes, with only one patient displaying a deletion in BCOR.

Tab 1. Total MACs samples

ID	PAX6	SOX2	OTX2	VSX2	GDF6	SHH	RAX	OTHER GENES
1	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
2	Neg	Neg	Neg	Neg	Neg	Neg	Neg	SMC3 R521Q ex16
3	Neg	Neg	Neg	Neg	Neg	Neg	c.290-4 C>A	Neg
4	Neg	Neg	c.97+12 C>T	Neg	Neg	Neg	Neg	Neg
5	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
6	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
7	Neg	Neg	Neg	Neg	Neg	Neg	Neg	BMP4 R211Q ex4
8	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
9	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
10	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
11	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
12	Neg	Neg	Neg	Neg	Neg	Neg	Neg	STRA6 Y18Tfs*54; F654X

13	Neg	Neg	Neg	Neg	Neg	Neg	Neg	STRA6 c.928-2 A>G
14	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
15	Neg	Neg	Neg	c.244delG p.L84S fs*57	Neg	Neg	Neg	Neg
16	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
17	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
18	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
19	Neg	Neg	Neg	Neg	c.1304C>T p.A435V	Neg	Neg	Neg
20	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
21	Neg	Neg	Neg	Neg	Neg	Neg	Neg	BCOR c.5041C>T p.R1681C
22	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
23	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
24	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
25	Neg	Neg	Neg	Neg	Neg	Neg	Neg	BMP4 c.979 G>A p.V327M
26	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
27	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
28	Neg	Neg	c.97+12 C>T	Neg		Neg	Neg	Neg
29	Neg	Neg	Neg	Neg	c.322 G>A p.A108T	Neg	Neg	Neg
30	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
31	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
32	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
33	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
34	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
35	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
36	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
37	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
38	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
39	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
40	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg
41	Neg	Neg	Neg	Neg	Neg	Neg	Neg	Neg

Tab.2. Variants found in MACs genes

ID	GENE	EXON	NUCLEOTIDE CHANGE	AMINOACID CHANGE	MUTATION TYPE
2	SMC3	16	c.1562 G>A	R521Q	Missense
3	RAX	Intr. 1	c.290-4C>A		Splicing Mutation?
4	OTX2	Intr. 3b	c.97+12 C>T		Splicing Mutation?
7	BMP4	4	c.632 G>A	R211Q	Missense
12	STRA6		c.52delT c.1961_1962	Y18Tfs*54 F654X	Deletion Nonsense
13	STRA6	Intr. 14	c.928-2 A>G		Splicing Mutation
15	VSX2	1	c.244 del G	L84S fs*57	Frameshift deletion
19	GDF6	2	c.1304 C>T	A435V	Missense
21	BCOR	16	c.5041 C>T	R1681C	Missense
25	BMP4	5	c.979 G>A	V327M	Missense
28	OTX2	Intr. 3b	c.97+12 C>T		Splicing Mutation?
29	GDF6	1	c.322 G>A	A108T	Missense

Detailed variants identified in MAC patients:

Missense mutations

Six missense mutations (46%) have been found in patients with anophthalmia and microphthalmia (A/M) diagnosis. Two mutations in the GDF6 gene, c.322G>A (p.A108T) and c.1304C>T (p.A435V), fall respectively into the N-terminal and C-terminal domains of the protein. For both variants, the small chemical-physical differences present in the aminoacids do not lead to drastic functional changes in the protein. The 4 remaining missense variants reported have been detected in additional genes of the MAC panel. The c.1562G>A (p.R521Q) mutation in the SMC3 gene may compromise the functionality of the protein as it falls into a functional region of the N-terminal domain (RecF/RecN/SMC domain). SMC3 gene is normally associated to a Cornelia-like phenotype. The variant c.5041C>T (p.R1681C) in the BCOR gene could

cause a functionality impairment of the polypeptidic chain, due to chemical-physical differences between the two aminoacids. The c.979G>A (p.V327M) mutation in the BMP4 gene falls into the C-terminal domain and is reported as deleterious or potentially pathogenetic in pathogenicity prediction databases.

Splicing mutations

The c.928-2A>G variant present in the STRA6 gene falls into the intron 14 acceptor splicing site. This change could lead to an abnormal splicing mechanism of the mRNA. The other two intronic variants detected, c.290-4C>A in the RAX gene and c.97+12C>T in the OTX2 gene, are nucleotide changes within the splicing regions, but their actual pathogenetic function is not known.

Frameshift deletions

The c.244delG (p.L84Sfs*57) deletion mutation in the VSX2 gene creates a reading frameshift from codon L84, resulting in a STOP codon 56 aminoacids downstream. Another deletion resulting in a frameshift, in particular a thymine at position c.52, has been found in patient N° 12 in the STRA6 gene.

Nonsense mutation

The only nonsense mutation found among the MAC genes patients has been detected in the STRA6 gene in patient N° 12. The variant is present in position c.1961_1962 and determines the modification of the phenylalanine codon with a stop codon.

Tab. 3 Total samples with CHARGE Syndrome

ID	CHD7
1	Neg
2	Neg
3	c.5241C>G p.Y1747X
4	Del ex1
5	c.5405-17G>A
6	Neg
7	c.2096G>A p.S699N
8	c.5625T>C p.P1875P
9	c.8994+35 T>C 3'UTR
10	Neg
11	Neg
12	Neg
13	c.7282C>T p.R2428X
14	c.2238+1G>A
15	Neg
16	Neg
17	c.6103+8C>T c.3765 G>A, p.P1255P
18	c.307T>A, p.S103T c.7682G>A, p.G2561E
19	c.3004C>T, p.Q1002X
20	c.6357C>A, p.D2119E
21	c.6111C>T, p.P2037P
22	Neg
23	Neg
24	c.5293-5313delGACTCAAGGTTAGTGCGAGCT
25	c.197delT c.200 T>A, p.T67K
26	Neg
27	c.1480 C>T, p.R494X
28	Neg
29	Neg
30	ex23-4C>T
31	c.307T>A, p.S103T
32	c.6850 C>T, p.R2284X
33	Neg
34	Neg

Tab. 4 Variants found in CHD7 gene

ID	GENE	EXON	NUCLEOTIDE CHANGE	AMINOACID CHANGE	MUTATION TYPE
3	CHD7	24	c.5241C>G	Y1747X	Nonsense
4	CHD7	1			Deletion
5	CHD7	Intr. 25	c.5405-17 G>A		Splicing Muatation?
7	CHD7	3	c.2096 G>A	S699N	Splicing Muation
8	CHD7	28	c.5625 T>C	P1875P	Synonymous
9	CHD7	Intr. 38	c.8994+35 T>C 3'UTR		Splicing Mutation
13	CHD7	34	c.7282 C>T	R2428X	Nonsense
14	CHD7	Intr. 4	c.2238+1G>A		Splicing Mutation
17	CHD7	Intr. 30	c.6103+8C>T		Splicing Mutation
		15	c.3765 G>A	P1255P	Synonymous
18	CHD7	35	c.7682 G>A	G2561E	Missense
		2	c.307 T>A	S103T	Missense
19	CHD7	12	c.3004 C>T	Q1002X	Nonsense
20	CHD7	31	c.6357 C>A	D2119E	Missense
21	CHD7	31	c.6111 C>T	P2037P	Synonymous
24	CHD7	24	c.5293-5300 +13 del		Deletion frameshift
25	CHD7	2	c.197delT	p.L66Rfs*17	Deletion
			c.200 T>A	T67K	Missense
27	CHD7	2	c.1480 C>T	R494X	Nonsense
30	CHD7	Intr. 22	c.5051 -4C>T		Splicing Mutation
31	CHD7	2	c.307 T>A	S103T	Missense
32	CHD7	32	c.6850 C>T	R2284X	Nonsense

Detailed variants identified in CHARGE patients:

Missense mutations

Among the 5 missense variants (22%) in the CHD7 gene, the detected variant c.307T>A (p.S103T) in two patients is potentially deleterious for the functionality of the protein according to the SIFT and PolyPhen genomic databases.

Stop mutations

Stop (nonsense) mutations in our study represent the 22% of the total mutations in the CHD7 gene and are located in different exons (2, 12, 24, 32, 34). The nucleotide changes leading to the formation of STOP codons could cause transcript degradation to avoid the formation of truncated proteins trough the mechanism of Nonsense Mediated Decay (NMD). The nonsense-mediated mRNA decay (NMD) pathway is well known as a translation-coupled quality control system that recognizes and degrades aberrant mRNAs with truncated open reading frames (ORF) due to the presence of a premature termination codon (PTC) [Schweingruber C. et al.; 2013]

.Splicing mutations

Five variants have been reported in non-coding regions, specifically involved in the splicing mechanism, in the CHD7 gene. In particular, the c.2238+1G>A variant in intron 4, could cause an anomalous mRNA formation. Also, the c.2096G>A (p.S699N) nucleotide change in the coding portion of exon 3 involves a specific sequence implicated in splicing.

Deletions

The large 13bp deletion (c.5293-5300+13del) found in patient N°24 involves the exon-intron 24 boundary, eliminating the whole region involved in splicing. A second deletion, c.197delT (p.L66Rfs*17), was found in exon 2 in patient N°25; while a third deletion has been detected in a non-coding, but potentially regulatory, region exon 1 of the CHD7 gene.

Synonymous mutations

Although synonymous mutations do not cause aminoacid changes, the 3 synonymous mutations detected, c.5625T>C (p.P1875P), c.3765G>A (p.P1255P) and c.6111C>T (p.P2037P), could compromise the functionality of the splicing mechanism due to the specific nucleotide changes involved.

3'UTR variant

The c.8994+35T>C nucleotide change in the 3'UTR of the CHD7 gene could cause an anomalous maturation mechanism of mRNA, as it involves a sequence dedicated to the recognition of specific proteins that constitute the complex spliceosome machinery.

Tab. 5 Total samples with Cornelia De Lange syndrome

ID	NIPBL	SMC3	SMC1A	HDAC8	RAD21	OTHER GENES
1	Neg	Neg	Neg	Neg	Neg	Neg
2	Neg	Neg	Neg	Neg	Neg	Neg
3	Neg	Neg	Neg	Neg	Neg	Neg
4	c.535G>A p.A179T	Neg	Neg	Neg		Neg
5	Neg	Neg	Neg	Neg	c.1635delA p.G547fs*65	Neg
6	Neg	Neg	Neg	Neg	Neg	Neg
7	Neg	Neg	Neg	Neg	Neg	Neg
8	Neg	Neg	Neg	Neg	Neg	Neg
9	c.6109-3T>C	Neg	Neg	Neg	Neg	Neg
10	Neg	Neg	Neg	Neg	Neg	Neg
11	Neg	Neg	Neg	Neg	Neg	Neg
12	Neg	Neg	Neg	Neg	Neg	Neg
13	Neg	Neg	Neg	Neg	Neg	Neg
14	Neg	Neg	Neg	Neg	Neg	Neg
15	c.6620 T>C p.M2207T	Neg	Neg	Neg	Neg	Neg
16	Neg	Neg	Neg	Neg	Neg	Neg
17	Neg	Neg	Neg	Neg	Neg	neg
18	Neg	Neg	Neg	Neg	Neg	Neg
19	Neg	Neg	Neg	Neg	Neg	Neg
20	Neg	Neg	Neg	Neg	Neg	Neg
21	c.781T>G p.S261A	Neg	Neg	Neg	Neg	Neg
22	Neg	Neg	Neg	Neg	Neg	Neg
23	Neg	Neg	Neg	Neg	Neg	Neg
24	Neg	Neg	Neg	Neg	Neg	Neg
25	Neg	Neg	Neg	Neg	Neg	Neg
26	Neg	Neg	Neg	Neg	Neg	Neg
27	Neg	Neg	Neg	Neg	Neg	Neg
28	Neg	Neg	Neg	Neg	Neg	Neg
29	c.240dupA p.D81fs*2	Neg	Neg	Neg	Neg	Neg
30	Neg	Neg	Neg	Neg	Neg	Neg
31	Neg	Neg	Neg	Neg	Neg	Neg
32	Neg	Neg	Neg	Neg	Neg	Neg
33	Neg	Neg	Neg	Neg	Neg	Neg
34	Neg	Neg	Neg	Neg	Neg	Neg
35	Neg	Neg	Neg	Neg	Neg	BCOR c.190A>G p.I64V
36	Neg	Neg	Neg	Neg	Neg	Neg
37	Neg	Neg	Neg	Neg	Neg	Neg

Tab. 6 Variants found in Cornelia de Lange syndrome

ID	GENE	EXON	NUCLEOTIDE CHANGE	AMINOACID CHANGE	MUTATION TYPE
4	NIPBL	2	c.535G>A	A179T	Missense
5	RAD21	13	c.1635delA	G547fs*65	Deletion frameshift
9	NIPBL	Intr. 34	c.6109-3 T>C		
15	NIPBL	39	c.6620 T>C	M2207T	Missense
21	NIPBL	8	c.781T>G	S261A	Missense
29	NIPBL	4	c.240dupA	D81fs*2	Insertion frameshift
35	BCOR	5	c.190A>G	I64V	Missense

Detailed variants identified in Cornelia de Lange patients:

Missense mutations

Four missense mutations (57%) have been found in patients with Cornelia De Lange, 3 of which are present in the NIPBL gene. In particular, the c.6620T>C (p.M2207T) mutation in the ARM like fold domain of the protein is reported as a possible causative mutation for Cornelia de Lange in pathogenicity prediction databases. The 4th missense variant reported, c.190A>G (p.I64V), was found in the BCOR gene, which is not directly involved in the syndrome but is included in the panel.

Deletions

The only deletion found is in the RAD21 gene. This deletion, c.1635delA (p.G547fs*65), determines a frameshift that leads to the formation of a STOP codon 64 aminoacids downstream.

Insertion

The c.240dupA (p.D81fs*2) insertion mutation found in exon 4 of the NIPBL gene is a duplication of an adenine causing a frameshift, which results in the formation of a stop mutation in the following codon.

Intronic variant

The c.6109-3T>C intronic variant detected in the NIPBL gene is located in the splicing region of exon 35. In silico predictions studies performed through pathogenicity prediction databases gave it as a benign variant.

Tab. 7 Total samples with Axenfeld-Rieger syndrome

ID	PITX2	FOXC1
1	Neg	775C>T p.L86F
2	c.355 C>T p.Q119X	Neg
3	Neg	Neg
4	Neg	Neg
5	Neg	Neg
6	Neg	Neg
7	Neg	Neg
8	Neg	Neg

Tab. 8 Variants found in Axenfeld-Rieger syndrome

ID	GENE	EXON	NUCLEOTIDE CHANGE	AMINO ACID CHANGE	MUTATION TYPE
1	PITX2	6	c.355 C>T	Q119X	Nonsense
2	FOXC1	1	c.256 C>T	L86F	Missense

The two mutations found in patients with Axenfeld-Rieger syndrome involve both genes associated to the disease. The c.355C>T (p.Q119X) mutation in the PITX2 gene introduces a stop codon in exon 6. The c.256C>T (p.L86F) mutation in the FOXC1 gene is a missense mutation which, from in silico prediction studies, appears as a causative mutation of the disease as it falls into the functional fork head domain, necessary for DNA binding. In the 8 patients analyzed for renal papilloma syndrome, no PAX2 variants have been detected, while among the 6 patients studied for Lenz microphthalmia, one patient presented a deletion in the BCOR gene.

5. DISCUSSION

This work is based on the development of a new molecular method that allows the analysis of diseases that involve many genes. The introduction of Next Generation Sequencing has been useful for the analysis of microphthalmia, anophthalmia, coloboma and associated syndromes, as it gives the possibility to evaluate large genomic regions reducing costs and working times. Actually, the possibility to obtain a a lot of information on the mutational state of the genes involved in eye development in relatively short timings, the diagnosis for A/M is definitely more exhaustive. Genotype-phenotype correlation studies can now be taken into account in the clinical evaluation. One of the main advances with NGS, concerning the molecular diagnosis, is the possibility to create panel of genes specific for different pathologies. In this view, molecular diagnosis of MACs became possible after the identification, by searching the literature, of the many genes involved in the pathogenic pathway. Following this, the use of *DesignStudio*, provided by Illumina, enabled the development of an appropriate panel comprising the genes of interest. The software helps the operator in the choice and elaboration of the amplicons to sequence. With the software help several specific probes encompassing the target regions thus designed. The sequencing method used by the NextSeq 500 system (Illumina) is the Sequencing By Synthesis (SBS), with the use of the Nextera Rapid Capture Custom Enrichment kit. Among the parameters evaluated by the software directly installed on the instrument, the Quality Score, which represents the probability that a base is correctly called, and the percentage of "Clusters Passing Filter", representing the percentage of clusters generated that will be used for the sequencing, are the most important parameters. About 2 Gb of genomic information obtained from the experiment are available through BaseSpace and Variant Studio software; the raw sequencing data are displayed through the IGV software. Each variant found, compared to a genomic reference sequence, has been also evaluated by Sanger sequencing. The traditional method allowed us to confirm the validity of the data obtained with next generation sequencing. The amplicons with a poor reading quality on IGV, have also been analyzed with the classical Sanger sequencing. With the analysis softwares provided by Illumina for NextSeq and the online genomic databases, it was possible to assign a meaning to the many variants found, although in some cases purely

theoretical. In the present study, NGS allowed us to analyze several genes for a given patient and to identify variants both in genomic regions directly involved in the pathogenesis of anophthalmia, microphthalmia and coloboma, both in genes not directly implicated in syndromic or isolated MACs. This feature allows us to understand at a molecular level many of those phenotypes that are shaded, atypical and not perfectly defined in a specific syndrome. In this regard, the continuous interaction with the referring physicians has been fundamental in our study. The results of the molecular analyses have always been evaluated in team, in some instances confirming the clinical suspect, in others calling in question the clinical signs and reevaluating the patient and the family members at a clinical level.

Tab.9 Pathogenicity predictions in variants of MACs genes

GENE	MUTATION	*f*	SIFT	MUTATION TASTER
SMC3	R521Q	-	Harmful	-
BMP4	R211Q	-	-	-
GDF6	A435V	0%	Tolerated	Cause of disease
BCOR	R1681C	-	Harmful	Cause of disease
BMP4	V327M	-	Harmful	Cause of disease
GDF6	A108T	0%	Tolerated	Cause of disease

Mutations in GDF6 usually lead to the manifestation of coloboma forms and bone malformations during development. The A435V and A108T variants found in the present study, have a very low frequency in the global population and their pathogenicity was discordant after analysis with two prediction softwares. Once the variants have been found, they have been loaded on Alamut software (Interactive Biosoftware) to define a pathogenicity prediction. Alamut retrieves information on aminoacid substitution from two sites: SIFT (Sorting Tolerant From Intolerance) and Mutation Taster; they are both based on algorithms able to elaborate the aminoacid position within the polypeptide chain under study. The position of the mutated aminoacid is compared to various species in order to determine its conservation during evolution. The higher the conservation of that aminoacid, the higher will be its importance in the protein function [Kumar P. et al.; 2009]. From this assumption the pathogenicity of the mutation is determined in silico (Tab. 9).

The other missense mutations found in MAC patients, never described in the literature, are very rare and have potentially harmful effects on the phenotype. For some of these (SMC3 R521Q, BMP4 R211Q) it was possible to perform familiar studies and to follow the mutation segregation (Fig.6).

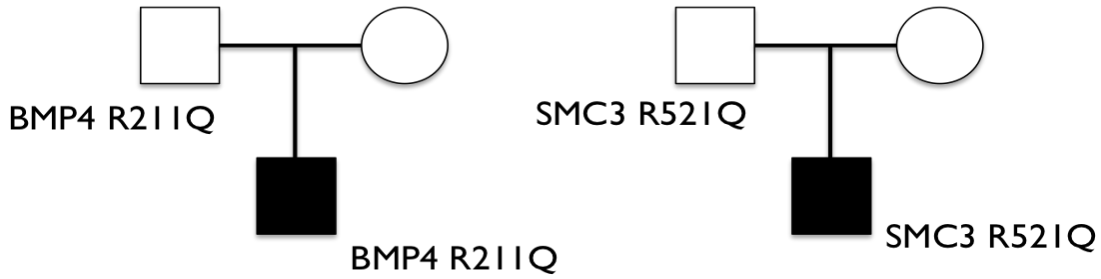

Fig. 6 Familiar studies for the variants R521Q in the SMC3 gene and R211Q in the BMP4 gene

Familial studies help us to elucidate if that variant has indeed a phenotypic effect, as the transmitted variant should have no meaning unless the carrier parent has also phenotypic features resembling the probands disease. Moreover with dominant pathologies, incomplete penetrance and variable expression have to be taken into account as possible mechanism. Family analysis also enables the formulation of the recurrence risk of the disease in subsequent generations.

In addition to the missense variants, accounting for 46% of mutations in our cohort of A/M patients, other variants include splicing mutations, deletions and intronic variants. Regarding the splicing mutations, the c.928-2A>G change in intron 14 of the STRA6 gene, could lead the exclusion of exon 15 from the transcript during the editing phase of mRNA.

Mutations in this gene are implicated in the Matthew-Wood syndrome (Syndromic microphthalmia, type 9), a rare clinical condition whose main features include severe anophthalmia or microphthalmia and hypoplasia or pulmonary aplasia. As this is a recessive condition, the second causative mutation was searched, although with no success. Regarding the STRA6 gene, we found two distinct pathogenic variants (Y18Tfs*54 and F654X) in a single patient (patient N°12). The first variant causes a frameshift that introduces a STOP codon 55 aminoacids downstream, while the second is a nonsense variant, also generating a premature interruption of the sequence. For this

patient, for which both parents were available, it was also possible to study the segregation of the mutations and to confirm the compound heterozygosity as each parent had one of the two mutations. During the genetic counseling, the couple was informed of their recurrence risk and of the possibility, in case of a future pregnancy, to monitor the state of the fetus at the molecular level.

Another deletion, L84Sfs*57 has been detected in the VSX2 gene. This deletion causes a frameshift which introduces a STOP codon 56 aminoacids downstream. The premature interruption of the mRNA could lead to mRNA degradation, to avoid the translation of an inactive truncated protein (NMD, nonsense mediated decay). Some variants reported in the study include intronic nucleotide changes. These intronic mutations have been investigated because they are near to coding sequences and are probably involved in the splicing process.

In our casuistry of 34 patients with CHARGE syndrome, 19 patients (56%) showed variants in the CHD7 gene (Fig. 7). The gene encodes for a protein involved in gene expression regulation and, when mutated, is responsible for a wide range of severe phenotypes. The variants found reflect a random and homogeneous distribution within the gene.

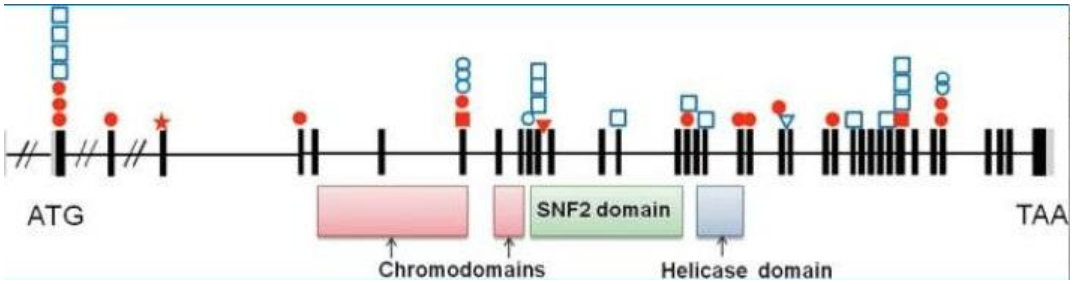

Fig.7 CHD7 gene

The mutations reported in the study include missense and nonsense variants, deletions, synonymous mutations and intronic variants involved in the spiicing process. The pathogenicity prediction for the missense mutations was made through SIFT and Mutation Taster, which gave us the possibility to compare modified aminoacids in the mutated sequence (Tab. 10).

Tab. 10. Pathogenicity prediction for the variants present in the CHD7 gene

GENE	**MUTATION**	*f*	**SIFT**	**MUTATION TASTER**
CHD7	S699N	-	Tolerated	Cause of disease
CHD7	G2561E	-	Tolerated	Cause of disease
CHD7	S103T	1%	Harmful	Cause of disease
CHD7	D2119E	-	Tolerated	Cause of disease
CHD7	T67K	-	Tolerated	Cause of disease

Most missense mutations, which accounted for 24% of the total mutations in CHD7, have not been described in the literature and their frequency is not predictable because they have never been found in the population. The discrepancy given by the softwares regarding pathogenicity prediction suggests the need of in vivo studies of the protein. The available data to study these variants include: Alamut, which is mainly used for in silico pathogenicity prediction and collects all information on the variant from the databases connected to it (SIFT, Mutation Taster, population frequency/1000 genomes, species conservation, scientific literature); PubMed and the scientific literature linked to the variant or to the gene involved; and familial studies. Variants introducing a STOP codon, hence providing an incomplete transcript, were found in exons 2, 12, 24, 32, 34. Even for the STOP mutations found we couldn't find any feedback from the literature. In two patients, we found variants interfering with the splicing process, such as the intronic variant c.2238+1G>A or the exon 3 change c.2096G>A (p.S699N), where the substitution of a guanine with an adenine in the coding region likely influences the splicing reaction. Also the c.8994+35T>C variant at the 3'UTR could be considered as potentially harmful for the general mechanism of mRNA formation. Although limited, our casuistry of CHD7 mutations showed a wide range of mutations. For example deletions in exons 2 and exon 24 were detected. In this latter case the deletion was located at position c.5293-5300+13 and involves the last nucleotides of exon 24 and the first of intron 24. This important deletion could result in a frameshift mutation affecting the reading frame in the following exon or have implications on the splicing process. The other deletion reported, c.197delT (p.L66Rfs*17), involves exon 2 in patient N°25, and results in a frameshift which introduces a STOP codon 17 aminoacids downstream. However, in this patient we also found the c.200T>A (p.T67K) variant. A mutational hotspot at this site for the gene could thus be suggested. In order to investigate the segregation of the two mutations, it was possible to carry out a familial study. Both parents were negative determining the *de novo* nature of the mutations. Hence, it was not possible to determine if the disposition of the variants was *in cis* or *in trans*. A 184 Kb deletion including both a part of the RAB2 gene and a part of the CHD7 gene was highlighted in patient N°4. The region involved in our gene of interest includes the non-coding exon 1 of CHD7, which is far from the ATG site, but probably important as a regulatory region.

It has been very interesting the study of synonymous mutations such as c.5625T>C (p.P1875P) (Fig.8), c.3765G>A (p.P1255P) and c.6111C>T (p.P2037P). Although their nucleotide changes do not cause an alteration in the translated aminoacid, they could cause changes in the splicing process. This was suggested by the Alamut software (Fig.9), which indicates the effect of a mutation on the success of splicing.

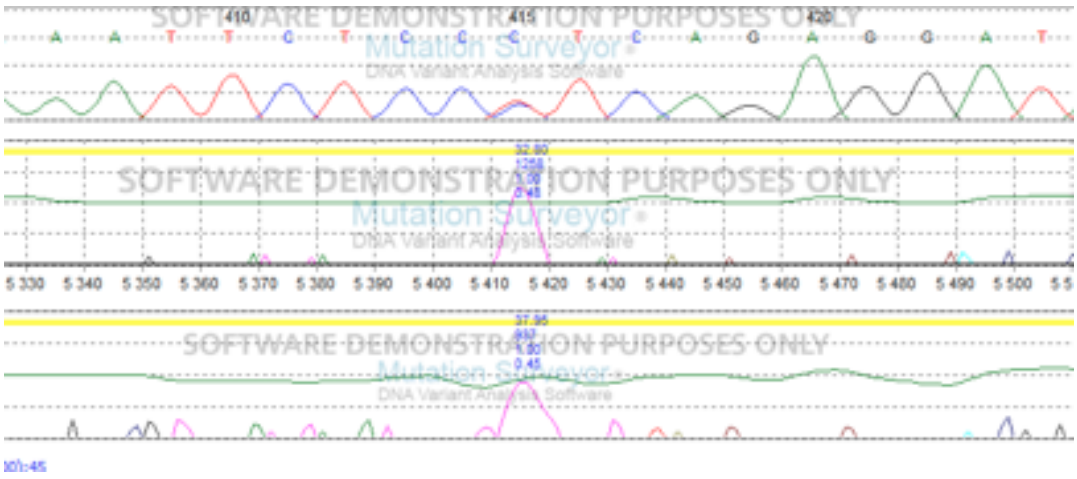

Fig. 8. Sequence of the c.5625T>C (p.P1875P) variant

Fig. 9. Splicing site variation due to c.5625T>C (p.P1875P) mutation

Our interest for these variants arose because they have never been found in the general population and are considered rare variants.

Among all patients analyzed with Cornelia De Lange syndrome, we found 7 variants (19%) (Tab. 11). Of these variants, 71% are located in NIPBL gene, in accordance with literature data [Deardorff M.A. et al., 2012], as NIPBL is the most frequently mutated gene in patients with Cornelia de Lange syndrome. Of the mutations found for these gene, 57% were missense mutations for which it was possible to determine a pathogenicity risk. In particular the c.781T>G (p.S261A) and c.6620T>C (p.M2207T) mutations in the NIPBL gene have been associated to the development of the disease by Mutation Taster and the latter (M2207T) has never been described in the literature. In addition, we found an adenine insertion, resulting in the c.240dupA (D81fs*2a) frameshift mutation, and an intronic variant, c.6109-3T>C, located in the splicing region. The c.1635delA (p.G547fs*65) mutation in the RAD21 gene was the only deletion found. This deletion causes a frameshift and introduces a STOP codon 64 aminoacids downstream. The BCOR gene has also been included in the panel of genes associated to A/M syndromic diseases. This allowed us to identify the c.190A>G (p.I64V) variant, reported as harmful according to in silico softwares and has never been described in the literature.

Tab. 11. Pathogenicity prediction of variants associated to Cornelia De Lange syndrome

GENE	MUTATION	*f*	SIFT	MUTATION TASTER
NIPBL	A179T	-	Tolerated	Polymorphism
NIPBL	M2207T	-	Harmful	Cause of disease
NIPBL	S261A	1%	Tolerated	Cause of disease
BCOR	I64V	-	Harmful	Polymorphism

A total of 8 patients have been analyzed for Axenfeld-Rieger syndrome. In a first patient we found the c.355C>T (p.Q119X) mutation in the PITX2 gene, determining the formation of a STOP codon in exon 6. In the second patient, the c.256C>T (p.L86F) variant was detected in the FOXC1 gene. This variant was reported as pathogenic by prediction softwares as it falls in a functional domain of the protein, the fork head domain, which is required to bind target regions of DNA (Tab.12).

Tab. 12. Pathogenicity prediction of the variant associated to Axenfeld-Rieger syndrome

GENE	MUTATION	*f*	SIFT	MUTATION TASTER
FOXC1	L86F	-	Harmful	Cause of disease

Respectively 8 and 6 patients were studied for PAX2 and BCOR genes. Of these, only one patient exhibits a deletion in the BCOR gene. The deletion has been detected by CGH array (Agilent platform) and spans 6.9 Mb (104 probes) from position 32.967.915 (Xp21.1) to position 39.913.316 (Xp11.4). This genomic region contains several genes including DMD (Dystrophin), XK (Kell blood group precursor), CYBB (Cytochrome b-245, beta polypeptide), RPGR (Retinitis pigmentosa GTPase regulator), OTC (Ornithine transcarbamylase), TSPAN7 (Tetraspanin 7) and BCOR (BCL6 corepressor). Many of the data found in this study highlight the importance to complement NGS technique with other methods, such as CGH array and mRNA studies. This is true for example for the detection of large deletions that would go unnoticed through Next Generation Sequencing, except for particular cases such as deletions of X chromosome in males, or is also true for understanding the path of mRNA transcripts with STOP mutations, for which mRNA studies are essential.

Surely the methodology we used to study MACs and their associated syndromes can help in the classification and definition of those phenotypes that are still difficult to classify, since all the potential variants that can be identified provide a wider understanding of the pathology. Also, such methodological approach with a broader spectrum of analysis can be very challenging and stimulating both for physicians and molecular geneticists, with the ultimate goal for both specialists to work in collaboration to find a link between the phenotypes described by the clinician and the many variants still to be identified by the geneticist.

6. CONCLUSIONS

Children with isolated or syndromic ocular malformations (MAC: microphthalmia, anophthalmia and coloboma) constitute a group of patients with a high degree of heterogeneity, both at phenotypic and genetic levels. Currently, national or international shared guidelines have yet to be defined and a phenotypic or genetic classification of the disease is still missing. Similarly to what happens for most rare diseases, the diagnostic and care management of children with ocular syndromes suffers heavily from the absence of strong scientific knowledge that would help in the development of shared guidelines, mainly because the phenotypic heterogeneity of these diseases is massive. The situation is slightly different for those patients with a more defined clinical phenotype, such as CHARGE, Cornelia de Lange and Axenfield-Rieger syndromes; however, we also have patients with Cornelia-Like phenotypes, and these patients will also show shaded clinical pictures compared to the classical forms.

In recent years, the study of isolated and syndromic MACs has had a remarkable development. The introduction of Next Generation Sequencing as a diagnostic tool is expanding enormously our knowledge on the molecular bases of pathogenesis. Our study also demonstrates that it is a crucial step to accompany functional studies to the potential amount of data that we are now able to acquire by this new technique.

Indeed, NGS is revolutionizing the approach of genetic studies, both for mendelian and complex diseases. Informatic tools can help in the definition of the pathogenicity prediction but will never replace an in vivo study. One of the goals for the future will be to create a multidisciplinary clinical pathway in oculistics, rare diseases and medical genetics to perform genotype-phenotype correlation studies with the aim of establishing a classification for isolated and syndromic MACs.

In addition, with the design of larger panels and the rapid evolution of exome sequencing for the identification of new genes involved in different pathologies, it will be possible to have deeper insights on the phenotype-genotype correlation of several diseases.

This will lead to a more comprehensive picture of the genetic and genomic bases of diseases that will eventually lead to a more specific diagnostic and care management.

7. BIBLIOGRAPHY

- Adler R, Canto-Soler MV Molecular mechanisms of optic vesicle development: complexities, ambiguities and controversies. Dev Biol. 2007 May 1; 305(1):1-13

- Albernaz VS, Castillo M, Hudgins PA, Mukherji SK. Imaging findings in patients with clinical anophthalmos. Am J Neuroradiol. 1997;18:555–561.

- Aydin B, Dilli D, Beken S, Zenciroglu A, Uzunalic N, Yuksekkaya P, Akyuz SG, Aydog O, Okumus N. Papillorenal syndrome with de novo reciprocal translocation t(2;15) (q31; q26). Genet Couns. 2013;24(2):201-5.

- Azuma N, Yamaguchi Y, Handa H, Hayakawa M, Kanai A, Yamada M Missense mutation in the alternative splice region of the PAX6 gene in eye anomalies. Am J Hum Genet. 1999 Sep; 65(3):656-63.

- Bailey TJ, El-Hodiri H, Zhang L, Shah R, Mathers PH, Jamrich M. Regulation of vertebrate eye development by Rx genes. Int J Dev Biol. 2004; 48(8-9):761-70

- Bardakjian T, Weiss A, Schneider A. Microphthalmia/Anophthalmia/Coloboma Spectrum. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJH, Bird TD, Ledbetter N, Mefford HC, Smith RJH, Stephens K, editors. GeneReviews®[Internet]. Seattle (WA): University of Washington, Seattle; 1993-2017. 2004 Jan 29 [updated 2015 Jul 9].

- Barishak YR. Embryology of the eye and its adnexae Dev Ophthalmol. 1992;24:1-142.

- Barisic I, Tokic V, Loane M, Bianchi F, Calzolari E, Garne E, Wellesley D, Dolk H; EUROCAT Working Group. Descriptive epidemiology of Cornelia de Lange syndrome in Europe. Am J Med Genet A. 2008 Jan 1;146A(1):51-9

- Bergman JE, Janssen N, Hoefsloot LH, Jongmans MC, Hofstra RM, van Ravenswaaij-Arts CM. CHD7 mutations and CHARGE syndrome: the clinical implications of an expanding phenotype. J Med Genet. 2011 May; 48(5):334-42.

- Blake KD, Prasad C. CHARGE syndrome Orphanet J Rare Dis. 2006 Sep 7; 1:34.

- Bower M, Salomon R, Allanson J, Antignac C, Benedicenti F, Benetti E, Binenbaum G, Jensen UB, Cochat P, DeCramer S, Dixon J, Drouin R, Falk MJ, Feret H, Gise R, Hunter A, Johnson K, Kumar R, Lavocat MP, Martin L, Morinière V, Mowat D, Murer L, Nguyen HT, Peretz-Amit G, Pierce E, Place E, Rodig N, Salerno A, Sastry S, Sato T, Sayer JA, Schaafsma GC, Shoemaker L,

Stockton DW, Tan WH, Tenconi R, Vanhille P, Vats A, Wang X, Warman B, Weleber RG, White SM, Wilson-Brackett C, Zand DJ, Eccles M, Schimmenti LA, Heidet L. Update of PAX2 mutations in renal coloboma syndrome and establishment of a locus specific database. Hum Mutat. 2012;33:457–66.

- Bower MA, Schimmenti LA, Eccles MR. Renal Coloboma Syndrome. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Mefford HC, Stephens K, Amemiya A, Ledbetter N, editors GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2017. 2007 Jun 08

- Boyle MI, Jespersgaard C, Brøndum-Nielsen K, Bisgaard AM, Tümer Z.Clin Cornelia de Lange syndrome. Genet. 2015 Jul; 88(1):1-12.

- Chow RL, Lang RA. Early eye development in vertebrates. Annu Rev Cell Dev Biol. 2001;17:255–96

- Cunliffe HE, McNoe LA, Ward TA, Devriendt K, Brunner HG, and Eccles MR. The prevalence of PAX2 mutations in patients with isolated colobomas or colobomas associated with urogenital anomalies. J Med Genet. 1998 Oct; 35(10): 806–812.

- Deardorff MA, Noon SE, Krantz ID. Cornelia de Lange syndrome. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJH, Bird TD, Ledbetter N, Mefford HC, Smith RJH, Stephens K, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2017. 2005 Sep 16

- Deardorff MA, Wilde JJ, Albrecht M, Dickinson E, Tennstedt S, Braunholz D, Monnich M, Yan Y, Xu W, Gil-Rodriguez MC, Clark D, Hakonarson H, et al. RAD21 mutations cause a human cohesinopathy. Am J Hum Genet. 2012b;90:1014–1027.

- Dong C, Yu B. Mutation surveyor: an in silico tool for sequencing analysis. Methods Mol Biol. 2011; 760:223-37.

- Feng L, Zhou D, Zhang Z, Liu Y, Yang Y. Exome sequencing identifies a de novo mutation in HDAC8 associated with Cornelia de Lange syndrome J Hum Genet. 2014 Sep; 59(9):536-9.

- Glenn TC Field guide to next-generation DNA sequencers. Mol Ecol Resour. 2011 Sep;11(5):759-69.

- Henderson RA, Williamson K, Cumming S, Clarke MP, Lynch SA, Hanson IM, FitzPatrick DR, Sisodiya S, van Heyningen V Inherited PAX6, NF1 and OTX2 mutations in a child with microphthalmia and aniridia. Eur J Hum Genet. 2007 Aug; 15(8):898-901.

- Hilton E, Johnston J, Whalen S, Okamoto N, Hatsukawa Y, Nishio J, Kohara H, Hirano Y, Mizuno S, Torii C, Kosaki K, Manouvrier S, Boute O, Perveen R, Law C, Moore A, Fitzpatrick D, Lemke J, Fellmann F, Debray FG, Dastot-Le-Moal F, Gerard M, Martin J, Bitoun P, Goossens M, Verloes A, Schinzel A, Bartholdi D, Bardakjian T, Hay B, Jenny K, Johnston K, Lyons M, Belmont JW, Biesecker LG, Giurgea I, Black G. BCOR analysis in patients with OFCD and Lenz microphthalmia syndromes, mental retardation with ocular anomalies, and cardiac laterality defects. Eur J Hum Genet. 2009 Oct;17(10):1325-35.

- http://www.interactive-biosoftware.com/doc/alamut-batch/1.7.0/EN/Alamut-Batch-1.7.0-User-Manual.pdf

- https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/nextseq/nextseq-500-system-guide-15046563-02.pdf

- https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_illumina_variantstudio_software.pdf

- https://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_nextera_rapid_capture_custom_enrichment.pdf

- Huisman S, Mulder PA, Redeker E, Bader I, Bisgaard AM, Brooks A, Cereda A, Cinca C, Clark D, Cormier-Daire V, Deardorff MA, Diderich K, Elting M, van Essen A, FitzPatrick D, Gervasini C, Gillessen-Kaesbach G, Girisha KM, Hilhorst-Hofstee Y, Hopman S, Horn D, Isrie M, Jansen S, Jespersgaard C, Kaiser FJ, Kaur M, Kleefstra T, Krantz ID, Lakeman P, Landlust A, Lessel D, Michot C, Moss J, Noon SE, Oliver C, Parenti I, Pie J, Ramos FJ, Rieubland C, Russo S, Selicorni A, Tümer Z, Vorstenbosch R, Wenger TL, van Balkom I, Piening S, Wierzba J, Hennekam RC. Phenotypes and genotypes in individuals with SMC1A variants. Am J Med Genet A. 2017 Aug;173(8):2108-2125.

- Kaur M, Mehta D, Noon SE, Deardorff MA, Zhang Z, Krantz ID. NIPBL expression levels in CdLS probands as a predictor of mutation type and phenotypic severity Am J Med Genet C Semin Med Genet. 2016 Jun; 172(2):163-70.

- Kelberman D, Islam L, Holder SE, Jacques TS, Calvas P, Hennekam RC, Nischal KK, Sowden JC Digenic inheritance of mutations in FOXC1 and PITX2 : correlating transcription factor function and Axenfeld-Riegerdisease severity. Hum Mutat. 2011 Oct; 32(10):1144-52.

- Kuhr, Werner G. and Monnig, Curtis A. 1992. Capillary Electrophoresis. Anal. Chem. 64:389R–406R.

- Kumar P., Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 2009a;4:1073–1081.

- Lalani SR, Safiullah AM, Fernbach SD, Harutyunyan KG, Thaller C, Peterson LE, McPherson JD, Gibbs RA, White LD, Hefner M, et al. Spectrum of CHD7 mutations in 110 individuals with CHARGE syndrome and genotype-phenotype correlation Am J Hum Genet. 2006 Feb; 78(2):303-14.

- Lewis R. Learning the 'SMART' way... results from a pilot study evaluating an interprofessional acute care study day. Nurse Educ Today. 2010 Jun;31(1):88-93

- Mannini L, Cucco F, Quarantotti V, Krantz ID, Musio A. Mutation spectrum and genotype-phenotype correlation in Cornelia de Lange syndrome.

- Morrison D, FitzPatrick D, Hanson I, Williamson K, van Heyningen V, Fleck B, Jones I, Chalmers J, Campbell H. National study of microphthalmia, anophthalmia, and coloboma (MAC) in Scotland: investigation of genetic aetiology. J Med Genet. 2002;39:16–22.

- Ng D. Lenz Microphthalmia Syndrome. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Mefford HC, Stephens K, Amemiya A, Ledbetter N, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2017. 2002 Jun 4.

- O'Keefe M, Webb M, Pashby RC, Wagman RD. Clinical anophthalmos. Br J Ophthalmol. 1987;71:635–638. doi: 10.1136/bjo.71.8.635.

- Opitz JM. The Brachmann-de Lange syndrome. Am J Med Genet. 1985 Sep;22(1):89-102. Review

- Pasutto F, Sticht H, Hammersen G, Gillessen-Kaesbach G, Fitzpatrick DR, Nurnberg G, Brasch F, Schirmer-Zimmermann H, Tolmie JL, Chitayat D, Houge G, Fernandez-Martinez L, Keating S, Mortier G, Hennekam RC, von der Wense A, Slavotinek A, Meinecke P, Bitoun P, Becker C, Nurnberg P, Reis A, Rauch A. Mutations in STRA6 cause a broad spectrum of malformations including anophthalmia, congenital heart defects, diaphragmatic hernia, alveolar capillary dysplasia, lung hypoplasia, and mental retardation. Am J Hum Genet. 2007;80:550–560.

- Pearce PM, Pitt DB. Six cases of de Lange's syndrome; parental consanguinity in two. Med J Aust. 1967 Mar 11;1(10):502-6

- Pisaneschi E, Sirleto P, Lepri FR, Genovese S, Dentici ML, Petrocchi S, Angioni A, Digilio MC, Dallapiccola B. CHARGE syndrome due to deletion of region upstream of CHD7 gene START codon. BMC Med Genet. 2015 Sep 3; 16:78.

- Plaisancie J, Calvas P, Chassaing N. Genetic Advances in Microphthalmia. J Pediatr Genet. 2016 Dec; 5(4):184-188.

- Ragge NK, Subak-Sharpe ID, Collin JR. A practical guide to the management of anophthalmia and microphthalmia. Eye. 2007; 21:1290-1300.

- Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E; Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Commitee. ACMG clinical laboratory standards for next-generation sequencing. Genet Med. 2013 Sep;15(9):733-47

- Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977 Dec;74(12):5463-7.

- Schneider A, Bardakjian T, Reis LM, Tyler RC, Semina EV Novel SOX2 mutations and genotype-phenotype correlation in anophthalmia and microphthalmia. Am J Med Genet A. 2009 Dec; 149A(12):2706-15.

- Schweingruber C, Rufener SC, Zünd D, Yamashita A, Mühlemann O. Nonsense-mediated mRNA decay - mechanisms of substrate mRNA recognition and degradation in mammalian cells. Biochim Biophys Acta. 2013 Jun-Jul; 1829(6-7):612-23.

- Shaham O, Menuchin Y, Farhy C, Ashery-Padan R. Pax6: a multi-level regulator of ocular development. Prog Retin Eye Res. 2012 Sep;31(5):351-76

- Song W, Hu X. The rare Axenfeld-Rieger syndrome with systemic anomalies: A case report and brief review of literature. Medicine (Baltimore). 2017 Aug;96(33):e7791

- Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration Brief Bioinform. 2013 Mar;14(2):178-92.

- Verma AS, Fitzpatrick DR. Anophthalmia and microphthalmia. Orphanet J Rare Dis. 2007 Nov 26;2:47

- Warburg M. An update on microphthalmos and coloboma. A brief survey of genetic disorders with microphthalmos and coloboma. Ophthalmic Paediatr Genet. 1991;12:57–63

- Williamson KA, FitzPatrick DR The genetic architecture of microphthalmia, anophthalmia and coloboma. Eur J Med Genet. 2014 Aug; 57(8):369-80.

- Zagozewski JL, Zhang Q, Eisenstat DD. Genetic regulation of vertebrate eye development. Clin Genet. 2014 Nov;86(5):453-60.